On Uniqueness and Weak Convergence of Solutions for the Stochastic Differential Equations of Nonlinear Filtering

by

Vladimir Lucic

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2001

0-612-60554-X

The University of Waterloo requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

Abstract

This thesis is a contribution to the theory of nonlinear filtering, and is concerned with two basic issues. The first of these deals with pathwise uniqueness and uniqueness in law for solutions of the measure-valued stochastic differential equations of nonlinear filtering, in the case where there is genuine dependence of the signal on the observation process. The second issue uses this uniqueness to establish convergence of nonlinear filters corresponding to a class of dynamical systems governed by sigularly perturbed stochastic differential equations, in which the perturbation is wide-band random process with certain ergodic properties.

Acknowledgements

I would like to thank my thesis advisor, Prof. Andrew Heunis, for inspiring my interest in Probability and being a constant source of knowledge, encouragement, and support along the way. His enthusiasm for research and commitment to excellence have had a profound effect on my academic career.

I would also like to thank the MITACS-PINTS research personnel, and in particular Prof. Michael Kouritzin, for their valuable comments on this work.

Finally, I would like to thank my wife Valentina for all the support she has given me over the years.

To my parents

Contents

Chapter 1

Introduction: Goals and Organization of Thesis

In this introductory chapter we give an informal summary of our main research goals, and then indicate the overall structure of the thesis.

1.1 Research Problem and Research Goals

This thesis is a contribution to the general theory of *nonlinear filtering*, a well-established branch of applied probability, which is important in many practical applications such as the inertial guidance of aircraft and space vehicles, and which also continues to offer interesting theoretical problems. Our interest here is less concerned with specific applications and is more focused on some of these theoretical problems. The present section provides a short and rather informal summary of the so-called nonlinear filtering problem, together with a brief discussion of the modeling issues involved. We will then indicate the main research problems of the thesis in the light of this summary.

In many problems of engineering one has a random process $\{X_t\}$, called the *signal*, which is defined on the probability space $(\Omega, \mathcal{F}, P)$ and takes values in some complete separable metric space E_1. For various reasons the signal cannot be observed directly, and hence is not precisely known to us. Together with the signal we also have available another process $\{Y_t\}$, called an *observation process*, defined on the same probability space $(\Omega, \mathcal{F}, P)$, which typically takes values in a finite dimensional Euclidean space $\mathbb{R}^r$, and which is somehow related to the signal. The observation process, in contrast to the signal, is precisely known to us, and in fact comprises all of our observed information about the signal. The so-called *nonlinear filtering problem* is to make "best " use of the information available in the observation process to ascertain "as much as possible" about the signal. To add some precision to this rather informal statement let us note that, in the circumstances, the most information that we can hope to ascertain about the signal X_t is its *conditional distribution* given the observation record $\{Y_s,\ 0 \leq s \leq t\}$. Thus, the goal of nonlinear filtering is to determine this conditional distribution at each instant t. Put another way, the goal of nonlinear filtering is to determine, at each instant t, the mapping $\pi_t : \Omega \to \mathcal{P}(E_1)$ (where — see the Glossary of Notation and Terminology in Appendix E — we use $\mathcal{P}(E_1)$ to denote the collection of all probability measures on the Borel σ-algebra of the metric space E_1) which gives the conditional distribution of X_t based on the observation $\{Y_s,\ 0 \leq s \leq t\}$, in the sense that

$$\pi_t f = \mathrm{E}[f(X_t) | Y_s,\ 0 \leq s \leq t] \ \ a.s. \tag{1.1.1}$$

for each t and each $f \in B(E_1)$. In this statement we have used (see the Glossary of Notation and Terminology in Appendix E) $B(E_1)$ to denote the set of all uniformly bounded Borel measurable mappings $f : E_1 \to \mathbb{R}$, and

$$\pi_t f := \int_{E_1} f(\xi)\, \pi_t(d\xi).$$

The $\mathcal{P}(E_1)$-valued process $\{\pi_t\}$ is usually called the *nonlinear filter of the signal* $\{X_t\}$ *corresponding to the observation process* $\{Y_t\}$. In order to deal with the problem of determining, or at least characterizing, this nonlinear filter $\{\pi_t\}$, we must postulate specific mathematical models for both the signal $\{X_t\}$ and for the relationship between the signal and observation process $\{Y_t\}$. The classical theory of nonlinear filtering, pioneered by Fujisaki, Kallianpur and Kunita [10], Zakai [39], and others, proceeds on the basis of these postulated models to derive recursive formulas, typically in the form of stochastic differential equations, which characterize the "probability measure-valued" process $\{\pi_t\}$, and which can, at least in principle, be "solved" for the process $\{\pi_t\}$ (these equations will be discussed in Chapter 2). It is in this sense that we attain our original goal of making best use of the information in the observation process to learn as much as we can about the signal process.

The mathematical models for the signal and for the relationship between the signal and observation process, can be formulated at various levels of generality, but it is always the case that the simpler the model the easier it is to "implement" the recursive formulas given to us by nonlinear filtering. We therefore want to use the very simplest models that are reasonably consistent with physical reality. A rather simple model that is very commonly used for the relation between the signal and observation process is of the form

$$\text{observation} = \text{some non-random function of signal} + \text{noise},$$

where the "noise" is Gaussian white noise. To make this model tractable by the methods of stochastic analysis, we re-write the preceding relation in "integrated" form, so that the observation process $\{Y_t\}$ is really modeled as an $\mathbb{R}^r$-valued process

$$Y_t := \int_0^t h(X_s)\, ds + W_t. \tag{1.1.2}$$

Here $h : E_1 \to \mathbb{R}^r$ is a Borel-measurable mapping (called the "sensor function") which is characteristic of the technical apparatus used to observe the signal, and $\{W_t\}$ is an $\mathbb{R}^r$-valued standard Wiener process on $(\Omega, \mathcal{F}, P)$ which models the "noise" that is an inevitable consequence of any attempt to observe the signal. Next, we consider mathematical models for the signal $\{X_t\}$ itself. A particularly important class of signals, very common in many applications, are those which take values in some finite-dimensional Euclidean space (i.e. $E_1 := \mathbb{R}^d$) and which represent the state of a *randomly perturbed dynamical system*. In this case a frequently used model for the signal $\{X_t\}$ is that it is the solution of an Itô stochastic differential equation (SDE) of the form

$$dX_t = b(X_t)\,dt + B(X_t)\,dW_t + c(X_t)\,dV_t \tag{1.1.3}$$

(see, e.g. Fujisaki, Kallianpur and Kunita [10]), in which the mappings $b : \mathbb{R}^d \to \mathbb{R}^d$, $B : \mathbb{R}^d \to \mathbb{R}^{d\times r}$, and and $c : \mathbb{R}^d \to \mathbb{R}^{d\times l}$ are at least locally bounded Borel measurable (and typically subject to other restrictions which ensure e.g. existence and some type of uniqueness), $\{V_t\}$ is an $\mathbb{R}^l$-valued standard Wiener process on the same probability space $(\Omega, \mathcal{F}, P)$ on which the $\mathbb{R}^r$-valued Wiener process $\{W_t\}$ in (1.1.2) is defined, and X_0, $\{V_t\}$, and $\{W_t\}$ are statistically independent. In (1.1.3), the mapping $b : \mathbb{R}^d \to \mathbb{R}^d$ accounts for the "aggregate" or "unperturbed" dynamics of the system, the term "$B(X_t)\,dW_t$" accounts for random perturbations that arise from possible "feedback" of the observation process $\{Y_t\}$ in (1.1.2) to the input of the dynamical system, and the term "$c(X_t)\,dV_t$" accounts for further perturbations that are internal to the dynamical system and statistically independent of the observation Wiener process $\{W_t\}$. In many applications there is actually no "feedback" of the observation process to the dynamics of the signal, in which case we can take $B \equiv 0$ in (1.1.3), so that the signal $\{X_t\}$ is modeled by the simpler relation

$$dX_t = b(X_t)\,dt + c(X_t)\,dV_t. \tag{1.1.4}$$

That said, it is nevertheless true that applications in which feedback is present are increasingly common (particularly in the inertial guidance of satellites and aircraft), and in this case the second term on the right hand side of (1.1.3) is an essential element of the model for the signal $\{X_t\}$. We note that the presence of this term introduces a dependence (or correlation) between the signal $\{X_t\}$ and the observation Wiener process $\{W_t\}$, which in turn often makes the analysis of the associated nonlinear filtering problem significantly more difficult than for the case where $B \equiv 0$ and $\{X_t\}$ and $\{W_t\}$ are consequently independent.

With the preceding discussion in mind we next formulate the rough outlines of the first of the two main research problems with which the present thesis will be concerned. Consider the nonlinear filtering problem in which the signal is modeled by (1.1.3) and the observation process is given by (1.1.2). The question of *uniqueness* of the solutions of the measure-valued SDEs which characterize the nonlinear filter $\{\pi_t\}$ has attracted considerable attention, and continues to offer interesting challenges. We are particularly interested in establishing *uniqueness in distribution* for the so-called *normalized filter equation* (this equation will be reviewed in Chapter 2). This question has already been studied by Szpirglas [36], who established uniqueness in distribution for the normalized filter equation in the special case where the signal $\{X_t\}$ is *independent* of the Wiener process $\{W_t\}$ in the observation equation (1.1.2). In the context of the preceding model this corresponds to taking $B \equiv 0$ in (1.1.3), or, equivalently, using (1.1.4) as the model of the signal. In the case where genuine feedback is present, so that $B \neq 0$ in (1.1.3), it turns out that the clever semigroup arguments of Szpirglas [36] no longer apply, and a fundamentally different approach is needed. We take up this question in Chapter 3, and establish uniqueness in law for the normalized filter equation when the signal is modeled by the SDE (1.1.3). The basic tools that we use are results on evolution equations (or forward equations) due to Kurtz [21].

The second research problem in this thesis may properly be viewed as an application of the uniqueness results outlined in the preceding paragraph, and is concerned with justifying the common practice of using Itô SDEs as models for the signal $\{X_t\}$. The main criticism of an Itô equation such as (1.1.4) is the use of a Wiener process $\{V_t\}$, with all of its drastic oversimplifications of physical reality, as the internal source of "randomness" in the model. We emphasize that this question extends well beyond the nonlinear filtering context that we are considering, and indeed makes sense in any situation where one wants to model a randomly perturbed dynamical system by means of Itô SDEs. In fact, this question goes to the very crux of what we truly understand by a "randomly perturbed dynamical system". This issue has been studied by Stratonovich [34], Khas'minskii [17], and Blankenship and Papanicolaou [5], who reach the conclusion that a more realistic and satisfactory notion of a "randomly perturbed dynamical system" is a singularly perturbed random ordinary differential equation (ODE) of the form

$$\frac{dX_t^\epsilon}{dt} = \frac{1}{\epsilon}F(X_t^\epsilon, Z_{t/\epsilon^2}) + G(X_t^\epsilon, Z_{t/\epsilon^2}). \tag{1.1.5}$$

The main elements of this model are

(i) a "weak mixing" process $\{Z_t\}$, taking values in some metric space S, such that Z_t converges in distribution to a probability measure $\bar{P}$ on the space S as $t \to \infty$.

(ii) mappings $F, G : \mathbb{R}^d \times S \to \mathbb{R}^d$ which are sufficiently well-behaved to ensure that (1.1.5) has a unique solution $\{X_t^\epsilon\}$ for each sample-path of the process $\{Z_t\}$.

(iii) a *small* parameter $\epsilon \in (0, 1]$.

In the model (1.1.5), the S-valued process $\{Z_t\}$ is to be regarded as the internal source of randomness, the vector field G appearing in the second term on the right side of (1.1.5) models the aggregate drift, and the vector field F in the first term on the right side of (1.1.5) models rapid local fluctuations around the paths determined by this drift. Notice that $F(x, Z_{s/\epsilon^2})$ and $F(x, Z_{t/\epsilon^2})$ are "almost independent" for fixed $x \in \mathbb{R}^d$ and $s \neq t$ (since $\epsilon > 0$ is small and $\{Z_t\}$ is "weak mixing"), which in turn ensures that the local fluctuations are indeed rapid provided that $\epsilon > 0$ is sufficiently small. We also want the local fluctuations contributed by the first term on the right of (1.1.5) to be essentially "aimless", without any overall drift, at least asymptotically for large t. To this end, the final element in the model given by (1.1.5) is that the condition

$$\int_S F(x, z)\bar{P}(dz) = 0, \quad \forall\, x \in \mathbb{R}^d, \tag{1.1.6}$$

be satisfied ($\bar{P}$ is the limiting probability measure in (i)). The main result linking the model (1.1.5) to solutions of the SDE (1.1.4), which has been established at varying levels of rigor and generality by Stratonovich [34], Khas'minskii [17], and Blankenship and Papanicolaou [5], is as follows: The $\mathbb{R}^d$-valued continuous process $\{X_t^\epsilon\}$ (given by (1.1.5)) converges in distribution to the $\mathbb{R}^d$-valued continuous solution $\{X_t\}$ of (1.1.4) as $\epsilon \to 0$, namely

$$\lim_{\epsilon \to 0+} \mathcal{L}(X^\epsilon) = \mathcal{L}(X) \quad \text{in} \quad \mathcal{P}(C_{\mathbb{R}^d}[0, \infty)), \tag{1.1.7}$$

subject to various technical conditions, and provided that the functions $b(\cdot)$ and $c(\cdot)$ in (1.1.4) are calculated in terms of the main elements of the model (1.1.5) using formulas established in [34], [17], and [5] (we will give these formulas in full detail in Chapter 4). The convergence (1.1.7) indicates that the SDE (1.1.4) is indeed a good approximation of the physically more realistic model (1.1.5) in the limit of exceedingly rapid fluctuations corresponding to very small ϵ, and really justifies the use of Itô SDEs as a model for randomly perturbed dynamical systems. The usefulness of this result derives from the fact that an Itô SDE is a far more tractable mathematical entity than the singularly perturbed random ODE (1.1.5), because the powerful stochastic calculus is available as a tool for dealing with SDEs such as (1.1.4).

With the preceding discussion in mind, we are finally able to outline the second of our two main research problems. Suppose that the signal is the $\mathbb{R}^d$-valued process modeled by the solution $\{X_t^\epsilon\}$ of the singularly perturbed ODE (1.1.5), with corresponding $\mathbb{R}^r$-valued observation

$$Y_t^\epsilon := \int_0^t h(X_s^\epsilon)\, ds + W_t, \tag{1.1.8}$$

for some $\mathbb{R}^r$-valued Wiener process $\{W_t\}$ which is independent of the signal $\{X_t^\epsilon\}$, and let $\{\pi_t^\epsilon\}$ be the nonlinear filter of the signal $\{X_t^\epsilon\}$ corresponding to the observation process $\{Y_t^\epsilon\}$. That is, $\{\pi_t^\epsilon\}$ is the $\mathcal{P}(\mathbb{R}^d)$-valued process such that

$$\pi_t^\epsilon f = \mathrm{E}[f(X_t^\epsilon)|Y_s^\epsilon,\, 0 \leq s \leq t] \;\; a.s. \tag{1.1.9}$$

for each $f \in B(\mathbb{R}^d)$ and t (c.f. (1.1.1)). Likewise, let $\{\bar{X}_t\}$ be the $\mathbb{R}^d$-valued signal which is modeled by the solution of the Itô SDE

$$d\bar{X}_t = b(\bar{X}_t)\, dt + c(\bar{X}_t)\, d\bar{V}_t, \tag{1.1.10}$$

for some $\mathbb{R}^l$-valued standard Wiener process $\{\bar{V}_t\}$ on the probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$, where the functions $b(\cdot)$ and $c(\cdot)$ in (1.1.9) have been determined in accordance with the formulas in Stratonovich [34], Khas'minskii [17] and Blankenship and Papanicolaou [5] to ensure that

$$\lim_{\epsilon \to 0+} \mathcal{L}(X^\epsilon) = \mathcal{L}(\bar{X}) \quad \text{in} \quad \mathcal{P}(C_{\mathbb{R}^d}[0,\infty)), \tag{1.1.11}$$

(c.f. (1.1.7)). We regard $\{\bar{X}_t\}$ as an $\mathbb{R}^d$-valued signal, with corresponding observation process $\{\bar{Y}_t\}$ defined by

$$\bar{Y}_t := \int_0^t h(\bar{X}_t)\, ds + \bar{W}_t, \tag{1.1.12}$$

for some $\mathbb{R}^r$-valued standard Wiener process $\{\bar{W}_t\}$ on $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$ that is independent of the signal $\{\bar{X}_t\}$. We now have a nonlinear filter $\{\bar{\pi}_t\}$ of the signal $\{\bar{X}_t\}$ corresponding to the observation process $\{\bar{Y}_t\}$, namely $\{\bar{\pi}_t\}$ is the $\mathcal{P}(\mathbb{R}^d)$-valued process with the property

$$\bar{\pi}_t f = \mathrm{E}[f(\bar{X}_t)|\bar{Y}_s,\ 0 \le s \le t] \ \ a.s. \tag{1.1.13}$$

for each $f \in B(\mathbb{R}^d)$ and t (c.f. (1.1.1)). It is then natural to enquire about the relation between the $\mathcal{P}(\mathbb{R}^d)$-valued processes $\{\pi_t^\epsilon\}$ and $\{\bar{\pi}_t\}$, and, in particular, in light of the convergence (1.1.11), does it necessarily follow that the nonlinear filter $\{\pi_t^\epsilon\}$ converges weakly to the "limiting" nonlinear filter $\{\bar{\pi}_t\}$ as $\epsilon \to 0$? In fact, the specific problem that we study is not quite the one just outlined, but has one additional element, namely the presence of "feedback" of the observation process as an input to the dynamics which model the signal. To account for this feedback we shall use, in place of (1.1.5), the singularly perturbed SDE

$$dX_t^\epsilon = \frac{1}{\epsilon} F(X_t^\epsilon, Z_{t/\epsilon^2})\, dt + G(X_t^\epsilon, Z_{t/\epsilon^2})\, dt + B(X_t^\epsilon)\, dW_t, \tag{1.1.14}$$

to model the signal $\{X_t^\epsilon\}$, with the corresponding observation process still being defined by (1.1.8). Here the mapping $B : \mathbb{R}^d \to \mathbb{R}^{d\times r}$ is not really an integral part of the signal model, but merely accounts for the effect of the Wiener process $\{W_t\}$, arising from feedback of the observation process $\{Y_t^\epsilon\}$ given by (1.1.8), on the dynamics of the signal $\{X_t^\epsilon\}$. It seems plausible that the convergence in (1.1.11) will continue to hold, provided that we modify the model (1.1.10) of the "limiting" signal $\{\bar{X}_t\}$ by adding an extra term to account for feedback of the observation process $\{\bar{Y}_t\}$ given by (1.1.12), namely

$$d\bar{X}_t = b(\bar{X}_t)\, dt + B(\bar{X}_t)\, d\bar{W}_t + c(\bar{X}_t)\, d\bar{V}_t. \tag{1.1.15}$$

The general question of convergence of the nonlinear filter $\{\pi_t^\epsilon\}$ to a limiting nonlinear filter $\{\bar{\pi}_t\}$, (as $\epsilon \to 0$) continues to make sense within this more general framework, and our second research problem is to establish this convergence. It turns out that the uniqueness results from Chapter 3 will play an essential role in dealing with this problem.

In summary, this thesis is concerned with two main research problems namely:

I: Establish uniqueness in law for the normalized filter equation corresponding to the signal model (1.1.3) and observation equation (1.1.2).

II: Establish convergence of the nonlinear filter $\{\pi_t^\epsilon\}$ defined by (1.1.9) (for signal $\{X_t^\epsilon\}$ given by (1.1.14) and observation $\{Y_t^\epsilon\}$ given by (1.1.8)), to the nonlinear filter $\{\bar{\pi}_t\}$ defined by (1.1.13) (for the signal $\{\bar{X}_t\}$ given by (1.1.15) and observation $\{\bar{Y}_t\}$ given by (1.1.12)).

1.2 Organization of the Thesis

In the present section we indicate the overall organization and structure of the thesis.

In Chapter 2 we give a background summary of some well-established ideas and results from the theory of nonlinear filtering, paying particular attention to the normalized and unnormalized filter equations, since these will be essential for the main developments of the thesis. We also establish some notation and terminology for later reference. This chapter does not contain any new results or ideas, but is included because there is little in the way of comprehensive and accessible literature on nonlinear filtering to which we can refer readers for the background that we need.

Chapter 3 takes up our first research problem, namely uniqueness in law for the normalized filter equation introduced in Chapter 2. We also compare the uniqueness results of this chapter with other results on uniqueness in the existing literature. The uniqueness established in this chapter will also be essential when we address the second research problem on convergence of nonlinear filters. Indeed, we will use the general method of martingale problems as the main device for establishing convergence of nonlinear filters, and this in turn involves well-posedness of the martingale problem corresponding to the "limiting" nonlinear filter $\{\bar{\pi}_t\}$. This well-posedness will result from the uniqueness in law established in Chapter 3.

Chapter 4 is concerned with some preliminaries which must be resolved before addressing the second research problem of the thesis. In particular, we establish the convergence (1.1.11) when $\{X_t^\epsilon\}$ is the solution of the singularly perturbed SDE (1.1.14) and $\{\bar{X}_t\}$ is the solution of the SDE (1.1.15). We shall use a method very similar to that of Blankenship and Papanicolaou [5], who show this convergence in the case where $\{X_t^\epsilon\}$ is given by the singularly perturbed ODE (1.1.5), that is $B \equiv 0$ in (1.1.14). Despite the strong similarity of our approach to that in [5], we give a complete proof of convergence, because dealing with the stochastic integral term on the right side of (1.1.14) is not entirely trivial, and also because the powerful theory of weak convergence of stochastic processes set forth in Ethier and Kurtz [8] (and summarized in Appendix D.1) considerably simplifies several of the arguments relative to those found in [5]. Most importantly, however, this chapter makes explicit the construction of so-called "perturbed test functions"; these are not only a necessary tool in establishing (1.1.11), but will also be essential for dealing with convergence of nonlinear filters.

In Chapter 5 we address the second problem of the thesis and show weak convergence of the nonlinear filter $\{\pi_t^\epsilon\}$ to the limiting nonlinear filter $\{\bar{\pi}_t\}$ as $\epsilon \to 0$. To this end, we shall introduce a martingale problem for the probability measure-valued solutions of the normalized filter equation introduced in Chapter 2, and then use the uniqueness result of Chapter 3 to show that the martingale problem for the limiting nonlinear filter $\{\bar{\pi}_t\}$ is well-posed. This will then enable us to apply a powerful convergence theorem of Bhatt and Karandikar [3] to establish convergence of the nonlinear filters. An essential role will be played by the perturbed test functions seen in Chapter 4, which will be essential in verifying some of the conditions of Bhatt and Karandikar's result. Finally, we compare the convergence results of this chapter with other studies on convergence of nonlinear filters in the established literature.

We end the thesis with a number of Appendices. Several of these are devoted to proofs of some of the more technical results needed in the various chapters. As a rule, we have tried to improve readability by placing in appendices proofs whose presence in the main text would otherwise obscure the overall pattern of ideas. We also draw the reader's attention to Appendix D, which has background on convergence of measures, the martingale problem, and some useful miscellaneous technical results.

Chapter 2

Background Summary of Nonlinear Filtering Theory

In this chapter we give a background summary of some well-established ideas and results from the theory of nonlinear filtering, paying particular attention to the normalized and unnormalized filter equations, since these will be essential for the main developments of the thesis. We also establish some notation and terminology for future reference. Our exposition is necessarily brief, all results being stated without proof, but we do provide full references where further details can be found.

2.1 The Nonlinear Filtering Problem

Let $T \in (0, \infty)$ be fixed and suppose that $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ is a complete filtered probability space on which we have a pair of $\{\mathcal{F}_t\}$-adapted stochastic processes $\{X_t,\ t \in [0,T]\}$ and $\{Y_t,\ t \in [0,T]\}$, such that $\{X_t\}$ takes value in a complete separable metric space E_1 and $\{Y_t\}$ is $\mathbb{R}^r$-valued. The indexing set $[0,T]$ is a finite "interval of interest". The process $\{X_t\}$ is called a *signal process* and represents the state of some physical system which is assumed not to be directly observable, while $\{Y_t\}$ is an *observation process* which is in some sense related to $\{X_t\}$. Within the nonlinear filtering framework, the natural filtration of $\{Y_t\}$ defined as

$$\mathcal{F}_t^Y := \sigma\{Y_s,\ 0 \le s \le t\} \vee \mathcal{N}(P), \quad t \in [0,T] \tag{2.1.1}$$

contains the only information available to us about the signal process $\{X_t\}$.

The goal of nonlinear filtering is twofold, namely

1. establish the existence and regularity properties of a "measure-valued" process $\{\pi_t,\ t \in [0,T]\}$, defined on $(\Omega, \mathcal{F}, P)$ and taking values in $\mathcal{P}(E_1)$, such that π_t is the conditional probability distribution of X_t given $\mathcal{F}_t^Y$ for each $t \in [0,T]$.

2. characterize in some useful sense (e.g. by "measure-valued" stochastic differential equations) the process $\{\pi_t\}$.

In order to attain these goals, we need to postulate some kind of dependence of the observation process $\{Y_t\}$ on the signal $\{X_t\}$, and we consider this next.

2.2 A Signal-Observation Model

Since some structure of sample paths of the signal process is necessary for most of the results that follow, we start by introducing the following condition.

Condition 2.2.1. *The signal process $\{X_t\}$ is a corlol E_1-valued $\{\mathcal{F}_t\}$-adapted process on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$.*

Remark 2.2.2. In the preceding statement we have used the acronym "corlol" for *continuous on the right, limits on the left* to indicate that the individual sample paths of $\{X_t\}$ are right-continuous at each $t \in [0,T)$ with left-limit in E_1 at each $t \in (0,T]$. A widely used alternative terminology is "càdlàg".

In many applications the observations can be assumed to have the form

observation = some non-random function of signal + noise,

where the "noise" is Gaussian white noise. To make this model tractable by the methods of stochastic calculus, we consider the "integrated version" of this relation, which leads to the following form of the observation process:

Condition 2.2.3. *The observation $\{Y_t\}$ is an $\mathbb{R}^r$-valued process on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ defined by*

$$Y_t = \int_0^t h(X_s)\,ds + W_t, \quad \forall t \in [0,T], \tag{2.2.2}$$

where

1. *$h : E_1 \to \mathbb{R}^r$ is a $\mathcal{B}(E_1)$-measurable function (often called the "sensor function") such that*

$$\mathrm{E}\left[\int_0^T |h(X_s)|^2\,ds\right] < \infty; \tag{2.2.3}$$

2. *$\{W_t,\ t \in [0,T]\}$ is an $\mathbb{R}^r$-valued $\{\mathcal{F}_t\}$-Wiener process on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$.*

The relation (2.2.3) is also known as the "finite-energy condition". It is a common standing assumption in nonlinear filtering, which is sufficient to ensure that the various ordinary and stochastic integrals that one encounters always exist are well behaved.

Remark 2.2.4. Observe that these conditions permit dependence of the signal process $\{X_t\}$ on the observation process $\{Y_t\}$, but are strong enough to ensure that

$$\sigma\{X_s, W_s,\ 0 \le s \le t\} \quad \text{and} \quad \sigma\{W_s - W_u,\ t < u < s \le T\}$$

are independent for each $t \in [0,T]$, since the σ-algebra on the left is included within $\mathcal{F}_t$, and $\{W_t\}$ is an $\{\mathcal{F}_t\}$-Wiener process.

Within this framework the two goals of nonlinear filtering defined in Section 2.1 have been completely attained, as we summarize in the next two sections.

2.3 An Existence Result of M. Yor

The first goal of nonlinear filtering is to establish the existence and regularity properties of a $\mathcal{P}(E_1)$-valued process $\{\pi_t\}$ which is the conditional probability of X_t given the observation σ-algebra $\mathcal{F}_t^Y$ (see (2.1.1)). The following result addressing that question is a special consequence of a general theorem of Yor [38].

Lemma 2.3.5. *Suppose that Condition 2.2.1 and Condition 2.2.3 hold. Then there exists a $\mathcal{P}(E_1)$-valued, corlol, and $\{\mathcal{F}_{t+}^Y\}$-optional process $\{\pi_t,\ t \in [0,T]\}$ on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ such that*

$$\pi_t f = \mathrm{E}[f(X_t)|\mathcal{F}_{t+}^Y] \quad a.s.$$

for each $t \in [0,T]$ and each $f \in B(E_1)$.

Remark 2.3.6. Since E_1 is a separable metric space the Borel σ-algebra $\mathcal{B}(E_1)$ is countably determined, and hence it follows that the process $\{\pi_t\}$ with the properties stated in Lemma 2.3.5 is uniquely determined to within indistinguishability

Remark 2.3.7. From now on we shall regard the $\mathcal{P}(E_1)$-valued corlol process $\{\pi_t\}$ given by Lemma 2.3.5 as the *nonlinear filter of the signal $\{X_t\}$ corresponding to the observation process $\{Y_t\}$* given by (2.2.2). Since (in view of conditional expectation as a projection operator) $\pi_t f$ minimizes the squared error loss for every $t \in [0,T]$, the $\mathcal{P}(E_1)$-valued process $\{\pi_t\}$ is also called the *optimal filter*. Notice that Lemma 2.3.5 is formulated in terms of conditioning not by $\mathcal{F}_t^Y$, but rather by its right continuous enlargement $\mathcal{F}_{t+}^Y$. From a practical viewpoint there is effectively no difference in conditioning by this latter σ-algebra, but $\{\mathcal{F}_{t+}^Y\}$ has the significant technical advantage of being a right-continuous filtration.

2.4 Normalized and Unnormalized Filter Equations

In this section we deal with the second goal of nonlinear filtering outlined in Section 2.1, and present the *Fujisaki-Kallianpur-Kunita equation* (also known as a *Kushner-Stratonovich equation* or the *normalized filter equation*), which is really a measure-valued stochastic differential equation characterizing the $\mathcal{P}(E_1)$-valued process $\{\pi_t\}$. We also present the closely related *Zakai equation*, or *unnormalized filter equation* which will be needed in Chapter 3.

The set-up is as follows: in addition to Condition 2.2.1 and Condition 2.2.3 we postulate a further condition which effectively defines a "dynamical model" for the signal $\{X_t\}$. This condition, which originates with Fujisaki, Kallianpur and Kunita [10], may appear to be rather abstract at first glance, but has the great merit that it encompasses an extremely broad variety of more specific signal models. Later in the thesis we shall avail ourselves of the generality of this condition.

Condition 2.4.8. *There exist operators $\mathcal{A} \subset \bar{C}(E_1) \times B(E_1)$ and $\mathcal{B}_k \subset \bar{C}(E_1) \times B(E_1)$, $1 \leq k \leq r$, having a common domain $\mathcal{D}(\mathcal{A}) \equiv \mathcal{D}(\mathcal{B}_k)$, $1 \leq k \leq r$ and such that*

1. *the signal process $\{X_t\}$ solves the martingale problem for $\mathcal{A}$, i.e. for every $\phi \in \mathcal{D}(\mathcal{A})$ the process*
$$M_t^{\phi} := \phi(X_t) - \int_0^t \mathcal{A}\phi(X_s)\, ds, \quad t \in [0, T] \tag{2.4.4}$$
is an $\{\mathcal{F}_t\}$-martingale;

2. *for every $\phi \in \mathcal{D}(\mathcal{A})$ we have*
$$\langle M^{\phi}, W^k \rangle_t = \int_0^t \mathcal{B}_k\phi(X_s)\, ds, \quad t \in [0, T],\ 1 \leq k \leq r. \tag{2.4.5}$$

Remark 2.4.9. The framework introduced in Condition 2.4.8 is general enough to include the vast majority of applications, in particular those where the signal process $\{X_t\}$ defined on some filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ takes values $E_1 := \mathbb{R}^d$, and is modeled by the solution of an Itô stochastic differential equation of the form

$$dX_t = b(X_t)\,dt + B(X_t)\,dW_t + c(X_t)\,dV_t, \quad t \in [0,T], \tag{2.4.6a}$$

for some locally bounded measurable functions $b : \mathbb{R}^d \to \mathbb{R}^d$, $B : \mathbb{R}^d \to \mathbb{R}^{d\times r}$, $c : \mathbb{R}^d \to \mathbb{R}^{d\times l}$, and an $\mathbb{R}^{r+l}$-dimensional $\{\mathcal{F}_t\}$-Wiener process $\{(W_t, V_t)\}$. The corresponding $\mathbb{R}^r$-dimensional observation process is defined by

$$Y_t = \int_0^t h(X_s)\,ds + W_t, \quad t \in [0,T], \tag{2.4.6b}$$

so that the signal process $\{X_t\}$ and the observation noise $\{W_t\}$ are correlated. Put

$$\begin{aligned}
&m(x) := \big[B(x) \;\; c(x)\big], \quad x \in \mathbb{R}^d;\\
&\mathcal{D}(\mathcal{A}) \equiv \mathcal{D}(\mathcal{B}_k) := C_c^\infty(\mathbb{R}^d);\\
&\mathcal{A}\phi(x) := \sum_{i=1}^d b^i(x)\partial_i\phi(x) + \frac{1}{2}\sum_{i,j=1}^d [m(x)m^T(x)]^{ij}\partial_i\partial_j\phi(x), \quad x \in \mathbb{R}^d,\ \phi \in C_c^\infty(\mathbb{R}^d);\\
&\mathcal{B}_k\phi(x) := \sum_{j=1}^d B^{jk}(x)\partial_j\phi(x), \quad x \in \mathbb{R}^d,\ \phi \in C_c^\infty(\mathbb{R}^d),\ 1 \le k \le r.
\end{aligned}$$

It is easy to check by Itô's formula (see Proposition 5.4.2 of Karatzas and Shreve [16]) that, for every $\phi \in C_c^\infty(\mathbb{R}^d)$, the process $\{M_t^\phi\}$ given by (2.4.4) is an $\{\mathcal{F}_t\}$-martingale, and (2.4.5) holds.

Remark 2.4.10. For every corlol $\{\mathcal{F}_{t+}^Y\}$-adapted $\mathcal{P}(E_1)$-valued process $\{\nu_t,\ t \in [0,T]\}$ on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ such that

$$\mathrm{E}\left[\int_0^T \nu_u|h|^2\,du\right] < \infty, \tag{2.4.7}$$

we define the $\mathbb{R}^r$-valued stochastic process $\{I_t^\nu,\ t \in [0,T]\}$ by

$$I_t^\nu := Y_t - \int_0^t \nu_u h\,du, \quad t \in [0,T]. \tag{2.4.8}$$

By taking $\{\nu_t\} \equiv \{\pi_t\}$ in (2.4.8) one gets the *innovations process* $\{I_t^\pi,\ t \in [0,T]\}$ which plays a central role in nonlinear filtering, and whose fundamental property is given by the following:

Lemma 2.4.11. *Suppose that Condition 2.2.1 and Condition 2.2.3 hold. Then* $\{(I_t^\pi, \mathcal{F}_{t+}^Y)\}$ *is an* $\mathbb{R}^r$*-valued Wiener process on* $(\Omega, \mathcal{F}, P)$.

The following fundamental result in nonlinear filtering theory, due to Fujisaki, Kallianpur and Kunita [10], provides a characterization of the nonlinear filter $\{\pi_t\}$.

Theorem 2.4.12. *Suppose that Condition 2.2.1, Condition 2.2.3, and Condition 2.4.8 hold. Then, for each* $\phi \in \mathcal{D}(\mathcal{A})$, *the corlol* $\mathcal{P}(E_1)$*-valued process* $\{\pi_t\}$ *furnished by Lemma 2.3.5 satisfies the stochastic integral relation: for every* $t \in [0,T]$

$$\pi_t\phi = \pi_0\phi + \int_0^t \pi_u(\mathcal{A}\phi)\,du + \sum_{k=1}^r \int_0^t [\pi_u(h^k\phi + \mathcal{B}_k\phi) - \pi_u(h^k)\pi_u(\phi)]\,d(I_u^\pi)^k, \tag{2.4.9}$$

subject to the initial condition

$$\pi_0\phi = \mathrm{E}[\phi(X_0)], \quad \forall\phi \in \mathcal{D}(\mathcal{A}). \tag{2.4.10}$$

Remark 2.4.13. Our statement of Theorem 2.4.12 follows Theorem VI.8.11 of Rogers and Williams [32]. The relation (2.4.9) may be regarded as a family of stochastic differential equations parametrized by $\phi \in \mathcal{D}(\mathcal{A})$, and is a basic result in nonlinear filtering. It is known variously as the Fujisaki-Kallianpur-Kunita equation or the *normalized filter equation.* We will use the latter terminology.

Remark 2.4.14. As an immediate consequence of (2.4.9) we see that $\{\pi_t\phi\}$ is an $\mathbb{R}$-valued *continuous* process for each $\phi \in \mathcal{D}(\mathcal{A})$. Thus, if $\mathcal{D}(\mathcal{A})$ is convergence determining (see Definition D.1.126), it follows that the $\mathcal{P}(E_1)$-valued process $\{\pi_t\}$ is continuous. We see, therefore, that in the presence of Condition 2.2.1 and Condition 2.2.3 the process $\{\pi_t\}$ is generally a *corlol* $\mathcal{P}(E_1)$-valued process (recall Lemma

2.3.5), but if, in addition, $\mathcal{D}(\mathcal{A})$ is convergence determining, then $\{\pi_t\}$ is necessarily a *continuous* $\mathcal{P}(E_1)$-valued process. For example, in the filtering problem considered in Remark 2.4.9 the operator $\mathcal{A}$ has domain $\mathcal{D}(\mathcal{A}) := C_c^\infty(\mathbb{R}^d)$, which is convergence determining (see Fact D.1.129). Therefore, the $\mathcal{P}(\mathbb{R}^d)$-valued nonlinear filter $\{\pi_t\}$, for the signal given by (2.4.6a) and observation given by (2.4.6b), has continuous sample paths.

Remark 2.4.15. The normalized filter equation (2.4.9) can be used to derive filters of genuine practical importance. For example, if the mappings $b(\cdot)$, $B(\cdot)$, $c(\cdot)$, and $h(\cdot)$ in (2.4.6a) and (2.2.2) are *linear mappings*, then (2.4.9) immediately yields the very widely used *Kalman-Bucy filter*. However, in general (2.4.9) is a nonlinear measure-valued SDE, and is consequently rather difficult to deal with. For this reason, we next introduce an alternative characterization of the nonlinear filter $\{\pi_t\}$ by means of a measure valued SDE which has a nice *bilinear* form, and which is considerably easier to work with than the normalized filter equation (2.4.9). To this end, define the $\mathcal{M}^+(E_1)$-valued process $\{\sigma_t,\ t \in [0,T]\}$ called the *unnormalized optimal filter* as follows: let

$$\eta_t := \exp\left\{\int_0^t \sum_{k=1}^r [\pi_s h^k]\, dY_s^k - \frac{1}{2}\int_0^t \sum_{k=1}^r [\pi_s h^k]^2\, ds\right\}, \quad t \in [0,T],$$

and put

$$\sigma_t(\Gamma) := \pi_t(\Gamma)\,\eta_t \quad t \in [0,T],\ \forall \Gamma \in \mathcal{B}(E_1). \tag{2.4.11}$$

It then follows that the nonlinear filter $\{\pi_t\}$ and the unnormalized optimal filter $\{\sigma_t\}$ are related via the formula

$$\pi_t\phi = \frac{\sigma_t\phi}{\sigma_t 1}, \quad t \in [0,T],\ \phi \in B(E_1).$$

Expanding $\{\sigma_t\phi\}$, $\phi \in \mathcal{D}(\mathcal{A})$, using (2.4.9) and Itô's formula, it is easy to verify that, for each $\phi \in \mathcal{D}(\mathcal{A}) \cup \{1\}$, the $\mathcal{M}^+(E_1)$-valued process $\{\sigma_t\}$ defined by (2.4.11) satisfies the stochastic integral relation

$$\sigma_t\phi = \sigma_0 f + \int_0^t \sigma_s(\mathcal{A}\phi)\,ds + \sum_{k=1}^{r} \int_0^t \sigma_s(h^k f + \mathcal{B}_k\phi)\,dY_s^k, \quad t \in [0,T]. \qquad (2.4.12)$$

This is called the *Zakai* or *Duncan-Mortensen-Zakai equation*, or simply the *unnormalized filter equation*. We will use the latter terminology. Clearly (2.4.12) is a considerably simpler relation than (2.4.9). Further simplification is possible if we can replace the ambient probability measure P with a related probability measure Q which makes the stochastic integrator $\{Y_t\}$ in (2.4.12) look like a Wiener process. To this end, define the measure Q on $(\Omega, \mathcal{F})$ by

$$Q(A) := \mathrm{E}\left[\frac{1}{\eta_T} I_A\right], \quad A \in \mathcal{F}. \qquad (2.4.13)$$

If one assumes that h is sufficiently bounded to ensure that Q is a *probability measure*, we can consider (2.4.12) under Q. In that case it follows from Girsanov's theorem that the observation process $\{Y_t\}$ is an $\{\mathcal{F}_t\}$-Wiener process. A sufficient condition ensuring that Q is a probability measure is, for example, the *Novikov condition* (see Corollary 3.5.13 of Karatzas and Shreve [16]):

$$\mathrm{E}\left[\exp\left(1/2 \int_0^T |h(X_s)|^2\,ds\right)\right] < \infty.$$

Remark 2.4.16. In the next chapter we shall take advantage of the simple structure of the unnormalized filter equation (2.4.12) in order to establish uniqueness in law for the normalized filter equation (2.4.9). As a by-product we shall also get pathwise uniqueness and uniqueness in law for the unnormalized filter equation.

Chapter 3

Uniqueness Results for the SDEs of Nonlinear Filtering

This chapter is devoted to the first of the research problems outlined in Section 1.1, namely the question of uniqueness in law for a specific instance of the normalized filter equation (2.4.9) introduced in Chapter 2. The main focus will be on proving uniqueness in law for the normalized filter equation which characterizes the nonlinear filter of the $\mathbb{R}^d$-valued signal $\{X_t\}$ given by the Itô SDE (2.4.6a) and corresponding to the $\mathbb{R}^r$-valued observation process $\{Y_t\}$ given by (2.4.6b). To this end, we will first establish pathwise uniqueness for the *unnormalized* filter equation, using in an essential way the structural simplicity of the unnormalized equation that was emphasized in Remark 2.4.15. Then we will use a well-known argument of Yamada and Watanabe [37] to conclude uniqueness in law for the *normalized* filter equation, as required.

We should point out that the whole issue of uniqueness for the nonlinear filter equations has received considerable attention, but none of the established results correspond to the type of uniqueness that we need, for reasons to be discussed later in the chapter. The results that we establish here may therefore have some interest outside the specific problem of convergence of nonlinear filters that is one of the principal goals of the thesis. The main results of this chapter are due to appear in [28].

3.1 Introduction

We begin this chapter with a general discussion of uniqueness for Itô SDEs, with particular emphasis on the ideas of Yamada and Watanabe [37] on weak solutions, pathwise uniqueness and uniqueness in law. These notions will then be adapted to the normalized and unnormalized filter equations.

Central to the study of a differential equation, ordinary or stochastic, is the question of the *uniqueness* of its solution(s). That is, given any two solutions of the differential equation subject to the same initial condition, does it necessarily follow that the solutions are "identical" in some sense. In the case of ODEs, there is a self-evident definition of uniqueness in which one declares uniqueness to hold for an ODE if any two solutions having common initial condition are equal for every subsequent value of the time parameter. It turns out that the situation is considerably more subtle in the case of Itô SDEs taking the form

$$dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dW_t, \quad t \in [0, T]. \tag{3.1.1}$$

In the course of studying this equation two distinct notions of uniqueness have emerged. Historically the first notion of uniqueness is so-called *pathwise uniqueness*, which somewhat parallels the notion of uniqueness for the deterministic (ODE) case. One says that **pathwise uniqueness** holds for (3.1.1) if, whenever $\{X_t^1\}$ and $\{X_t^2\}$ are two solutions on a common filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ corresponding to the same $\{\mathcal{F}_t\}$-Wiener process $\{W_t\}$, adapted to $\{\mathcal{F}_t\}$, and having

P-almost surely the same initial conditions, then $\{X_t^1\}$ and $\{X_t^2\}$ are necessarily indistinguishable.

The notion of pathwise uniqueness, although quite natural and intuitive, fails to provide a framework sufficiently rich to capture many relevant properties of (3.1.1). In particular, pathwise uniqueness is not always an appropriate notion when one is concerned with uniqueness of the *probability laws* of the solution of (3.1.1), a question that naturally arises in the study of the martingale problem associated with (3.1.1). This consideration leads to the notions of *weak solution* and *uniqueness in law* that focus on the distributional properties of a solution, rather than its path-related properties. Thus, one defines a **weak solution** of (3.1.1) as the pair $\{(X_t, W_t),\ (\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)\}$ where $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ is a complete filtered probability space, $\{W_t\}$ is an $\{\mathcal{F}_t\}$-Wiener process, and $\{X_t\}$ is a continuous $\{\mathcal{F}_t\}$-adapted process satisfying (3.1.1). It is worth noting that this definition specifies the probability space as a part of the solution, so that, in general, two weak solutions cannot be compared in the pathwise sense since they are generally defined on different probability spaces. A related notion of uniqueness, due to Yamada and Watanabe [37], uses the concept of a weak solution to capture distributional uniqueness of candidate solutions of (3.1.1): we say that **uniqueness in law** holds for (3.1.1) if, for any two weak solutions $\{(X_t, W_t),\ (\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)\}$ and $\{(\bar{X}_t, \bar{W}_t),\ (\bar{\Omega}, \bar{\mathcal{F}}, \{\bar{\mathcal{F}}_t\}, \bar{P})\}$ with X_0 and $\bar{X}_0$ having the same distribution, it follows that the processes $\{X_t\}$ and $\{\bar{X}_t\}$ are identically distributed.

Since the two definitions of uniqueness just introduced apparently bear little similarity to one another, it is natural to ask whether there is any relationship between them, and, in particular, whether one form of uniqueness implies the other. This turns out to be a decidely nontrivial question, to which the answer was provided by Yamada and Watanabe in their seminal paper [37] (for a more up-to date account see Section IV.1 of Ikeda and Watanabe [13] or Section IX.1 of Revuz and

Yor [30]). By constructing a common probability space for a pair of weak solutions for (3.1.1), Yamada and Watanabe establish a foundational result in the theory of SDEs, namely *pathwise uniqueness implies uniqueness in law.* We note in passing that a well-known counterexample due to H. Tanaka (see Karatzas and Shreve [16], Example 5.3.5) provides an SDE of the form (3.1.1) which has the property of uniqueness in law, but fails to have the property of pathwise uniqueness. Pathwise uniqueness is therefore a nontrivially stronger property than uniqueness in law.

The preceding ideas of Yamada and Watanabe, although developed in the context of Itô SDEs, in fact carry over to the normalized and unnormalized filter equations, as was demonstrated by Szpirglas [36]. The basic viewpoint adopted in [36] is to regard the normalized and unnormalized filter equations as entities quite separate from the original nonlinear filtering problem, for which one can formulate the notions of solution (or weak solution), pathwise uniqueness and uniqueness in law, by essentially adapting these concepts from the theory of Itô SDEs outlined at the beginning of this section. With these notions at hand, it is then established in [36] that pathwise uniqueness and uniqueness in law hold for both the normalized and unnormalized filter equations, in the case where the signal is a Markov process which is *independent of the Wiener process in the observation equation*, and the sensor function in the observation equation is uniformly bounded.

Our goal is to look at uniqueness for the SDEs of nonlinear filtering from a point of view very similar to that of Szpirglas [36], but for a nonlinear filtering problem in which there is genuine dependence of the signal on the Wiener process of the observation equation. In fact, we shall look at the specific nonlinear filtering problem in Remark 2.4.9, where the signal $\{X_t\}$ is an $\mathbb{R}^d$-valued process solving an equation of the form

$$dX_t = b(X_t)\,dt + B(X_t)\,dW_t + c(X_t)\,dV_t, \quad t \in [0,T], \tag{3.1.2}$$

the $\mathbb{R}^r$-valued observation process $\{Y_t\}$ is defined by

$$Y_t = W_t + \int_0^t h(X_s)\,ds, \tag{3.1.3}$$

and $\{(W_t, V_t)\}$ is a standard $\mathbb{R}^{r+l}$-valued Wiener process (precise conditions on the mappings $b(\cdot)$, $B(\cdot)$, $c(\cdot)$, and $h(\cdot)$ will be stated in Section 3.2).

Motivated by Szpirglas [36], in the following section we shall regard the normalized and unnormalized filter equations for this nonlinear filtering problem as measure-valued stochastic differential equations, defined quite independently of the filtering problem, and will formulate the notions of weak solution, pathwise uniqueness, and uniqueness in law for the filter equations. Our main result (see Theorem 3.3.34) establishes pathwise uniqueness for the unnormalized filter equation, together with uniqueness in law for the normalized and unnormalized filter equations, subject to reasonably general conditions on the mappings $b(\cdot)$, $B(\cdot)$, and $c(\cdot)$ in the signal equation (3.1.2), and a uniform boundedness condition on the sensor function $h(\cdot)$ in the observation equation (3.1.3). As will be seen from the discussion in Remark 3.3.35, the elegant semigroup ideas used in Szpirglas [36] do not extend to the filtering problem represented by (3.1.2) and (3.1.3), where the signal $\{X_t\}$ depends on the Wiener process $\{W_t\}$, and our approach necessarily involves a different method of proof.

3.2 Notions of Uniqueness for the SDEs of Nonlinear Filtering

We first recall the main features of the nonlinear filtering problem that was outlined in Remark 2.4.9, and which comprises the following elements:

E.1 A fixed interval of interest $[0, T]$, with $T \in (0, \infty)$.

E.2 Defined on a complete filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$, is an $\mathbb{R}^d$-valued continuous $\{\mathcal{F}_t\}$-adapted process $\{X_t,\ t \in [0,T]\}$ and an $\mathbb{R}^{r+l}$-valued $\{\mathcal{F}_t\}$-Wiener process $\{(W_t, V_t),\ t \in [0,T]\}$ such that (3.1.2) holds, where $b : \mathbb{R}^d \to \mathbb{R}^d$, $B : \mathbb{R}^d \to \mathbb{R}^{d\times r}$, and $c : \mathbb{R}^d \to \mathbb{R}^{d\times l}$ are Borel-measurable and locally bounded functions (that is, uniformly bounded over bounded subsets of $\mathbb{R}^d$).

E.3 an $\mathbb{R}^r$-valued *observation process* $\{Y_t,\ t \in [0,T]\}$ defined by (3.1.3), where $h : \mathbb{R}^d \to \mathbb{R}^r$ is Borel-measurable, with

$$\mathrm{E}\left[\sum_{k=1}^{r} \int_0^T |h^k(X_u)|^2\, du\right] < \infty. \tag{3.2.4}$$

Define the *observation filtration* $\{\mathcal{F}_t^Y,\ t \in [0,T]\}$ by

$$\mathcal{F}_t^Y := \sigma\{Y_u,\ u \in [0,t]\} \vee \mathcal{N}(P), \qquad \text{where} \qquad \mathcal{N}(P) := \{N \in \mathcal{F} :\ P(N) = 0\}. \tag{3.2.5}$$

From Lemma 2.3.5 there exists some $\mathcal{P}(\mathbb{R}^d)$-valued $\{\mathcal{F}_{t+}^Y\}$-optional process $\{\pi_t,\ t \in [0,T]\}$, called the *optimal filter*, which is defined on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ and satisfies

$$\pi_t\phi = \mathrm{E}[\phi(X_t)|\mathcal{F}_{t+}^Y] \quad a.s., \quad \forall t \in [0,T], \quad \forall \phi \in B(\mathbb{R}^d). \tag{3.2.6}$$

From (3.2.4) and Jensen's inequality we see that

$$\mathrm{E}\left[\sum_{k=1}^{r} \int_0^T [\pi_u(|h^k|)]^2\, du\right] < \infty,$$

and we can therefore define the $\mathbb{R}^r$-valued *innovations process* $\{I_t,\ t \in [0,T]\}$ by

$$I_t^k := Y_t^k - \int_0^t \pi_s h^k\, ds, \quad \forall\, t \in [0,T],\ k = 1, \ldots, r. \tag{3.2.7}$$

From Lemma 2.4.11 we know that $\{I_t,\ t \in [0,T]\}$ is an $\mathbb{R}^r$-valued $\{\mathcal{F}_{t+}^Y\}$-Wiener process, thus, since $\{I_t\}$ is continuous, it is necessarily $\{\mathcal{F}_t^Y\}$-adapted. Therefore, $\{I_t,\ t \in [0,T]\}$ is a $\{\mathcal{F}_t^Y\}$-Wiener process. Now define $m : \mathbb{R}^d \to \mathbb{R}^{d\times(r+l)}$ by

$$m(x) := \begin{bmatrix} B(x) & c(x) \end{bmatrix}, \quad \forall\, x \in \mathbb{R}^d,$$

and put

$$\mathcal{A}\phi(x) := \sum_{i=1}^{d} b^i(x)\partial_i\phi(x) + \frac{1}{2}\sum_{i,j=1}^{d}[m(x)m^T(x)]^{ij}\partial_i\partial_j\phi(x), \quad \forall x \in \mathbb{R}^d,\ \phi \in C^\infty(\mathbb{R}^d); \tag{3.2.8a}$$

$$\mathcal{B}_k\phi(x) := \sum_{i=1}^{d} B^{ik}(x)\partial_i\phi(x), \quad \forall x \in \mathbb{R}^d,\ \phi \in C^\infty(\mathbb{R}^d),\ k = 1, \dots, r. \tag{3.2.8b}$$

From Remark 2.4.9, for each $\phi \in C_c^\infty(\mathbb{R}^d)$ we know that

$$M_t^\phi := \phi(X_t) - \int_0^t \mathcal{A}\phi(X_s)\,ds, \quad t \in [0,T], \tag{3.2.9}$$

is an $\{\mathcal{F}_t\}$-martingale with

$$\langle M^\phi, W^k\rangle_t = \int_0^t \mathcal{B}_k\phi(X_u)\,du, \quad t \in [0,T],\ k = 1, \dots, r. \tag{3.2.10}$$

Taking $E_1 := \mathbb{R}^d$, we have now verified Conditions 2.2.1, 2.2.3, and 2.4.8, and hence from Theorem 2.4.12 we obtain

Theorem 3.2.17. *For the nonlinear filtering problem given by* **E.1**, **E.2**, *and* **E.3**, *the $\mathcal{P}(\mathbb{R}^d)$-valued optimal filter $\{\pi_t,\ t \in [0,T]\}$ satisfies the stochastic integral relation*

$$\pi_t\phi = \pi_0\phi + \int_0^t \pi_s(\mathcal{A}\phi)\,ds + \int_0^t \sum_{k=1}^{r}[\pi_s(h^k\phi + \mathcal{B}_k\phi) - (\pi_s h^k)(\pi_s\phi)]\,dI_s^k, \quad \forall\, 0 \le t \le T, \tag{3.2.11}$$

for each $\phi \in C_c^\infty(\mathbb{R}^d)$.

The relation (3.2.11) is the normalized filter equation for the filtering set-up of this chapter.

Remark 3.2.18. We have seen in Remark 2.4.14 that $\{\pi_t,\ t \in [0,T]\}$ is a *continuous* $\mathcal{P}(\mathbb{R}^d)$-valued process adapted to $\{\mathcal{F}_{t+}^Y\}$. Thus, it is also adapted to $\{\mathcal{F}_t^Y\}$, hence we can replace $\mathcal{F}_{t+}^Y$ in (3.2.6) by $\mathcal{F}_t^Y$.

Motivated by the notions of weak solution, pathwise uniqueness, and uniqueness in law introduced by Yamada and Watanabe [37] for Itô SDEs, Szpirglas [36] formulated the analogous concepts for a particular normalized filter equation (c.f. Szpirglas [36], Définition III.1, V.1, V.2). Here we do exactly the same thing for the normalized filter equation (3.2.11):

Definition 3.2.19. The pair $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{P}), (\tilde{\pi}_t, \tilde{I}_t)\}$ is a *weak solution* of the normalized filter equation when:

1. $(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{P})$ is a complete filtered probability space;
2. $\{\tilde{I}_t,\ t \in [0,T]\}$ is an $\mathbb{R}^r$-valued $\{\tilde{\mathcal{F}}_t\}$-Wiener process on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$;
3. $\{\tilde{\pi}_t,\ t \in [0,T]\}$ is a $\mathcal{P}(\mathbb{R}^d)$-valued continuous $\{\tilde{\mathcal{F}}_t\}$-adapted process such that

$$\tilde{P}\left(\int_0^T \sum_{k=1}^r [\tilde{\pi}_s|h^k|]^2\, ds < \infty\right) = 1, \tag{3.2.12}$$

and, for each $\phi \in C_c^\infty(\mathbb{R}^d)$, the following holds to within indistinguishability

$$\tilde{\pi}_t\phi = \tilde{\pi}_0\phi + \int_0^t \tilde{\pi}_s(\mathcal{A}\phi)\, ds + \sum_{k=1}^r \int_0^t [\tilde{\pi}_s(h^k\phi + \mathcal{B}_k\phi) - (\tilde{\pi}_s h^k)(\tilde{\pi}_s\phi)]\, d\tilde{I}_s^k,\ t \in [0,T]. \tag{3.2.13}$$

Remark 3.2.20. In view of Definition 3.2.19, it follows that $\{(\Omega, \mathcal{F}, \{\mathcal{F}_t^Y\}, P), (\pi_t, I_t)\}$ for $\{\mathcal{F}_t^Y,\ t \in [0,T]\}$, $\{\pi_t,\ t \in [0,T]\}$, and $\{I_t,\ t \in [0,T]\}$ defined by (3.2.5), (3.2.6), and (3.2.7) is a weak solution of the normalized filter equation.

Definition 3.2.21. The normalized filter equation has the property of *pathwise uniqueness* when the following holds: If $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{P}), (\tilde{\pi}_t^1, \tilde{I}_t)\}$ and $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{P}), (\tilde{\pi}_t^2, \tilde{I}_t)\}$ are weak solutions of the normalized filter equation with $\tilde{P}(\tilde{\pi}_0^1 = \tilde{\pi}_0^2) = 1$, then

$$\tilde{P}\left(\tilde{\pi}_t^1 = \tilde{\pi}_t^2 \quad \forall t \in [0,T]\right) = 1.$$

Definition 3.2.22. The normalized filter equation has the property of *uniqueness in joint law* when the following holds: If $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{P}), (\tilde{\pi}_t, \tilde{I}_t)\}$ and $\{(\bar{\Omega}, \bar{\mathcal{F}}, \{\bar{\mathcal{F}}_t\}, \bar{P}), (\bar{\pi}_t, \bar{I}_t)\}$ are weak solutions of the normalized filter equation with $\mathcal{L}_{\tilde{P}}(\tilde{\pi}_0) = \mathcal{L}_{\bar{P}}(\bar{\pi}_0)$, then the processes $\{(\tilde{\pi}_t, \tilde{I}_t),\ t \in [0,T]\}$ and $\{(\bar{\pi}_t, \bar{I}_t),\ t \in [0,T]\}$ have the same finite-dimensional distributions.

Remark 3.2.23. We next turn to the issue of formulating a notion of weak solution for an *unnormalized filter equation*, whose form should bear the same general relation to (3.2.13) as the Zakai equation (2.4.12) does to the Fujisaki-Kallianpur-Kunita equation (2.4.9). To get the form of this equation we essentially repeat the steps that led from (2.4.9) to (2.4.12), but now we work in the context of an *arbitrary* weak solution of the normalized filter equation, rather than the specific weak solution given by the optimal filter $\{\pi_t\}$ (recall Remark 3.2.20). To facilitate these steps, we will use the following notation: If $(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{P})$ is a complete filtered probability space, $\{\tilde{M}_t\}$ is a continuous $\{\tilde{\mathcal{F}}_t\}$-semimartingale, and $\{\tilde{\gamma}_t\}$ is a locally bounded $\{\tilde{\mathcal{F}}_t\}$-progressively measurable process, then $\tilde{\gamma} \bullet \tilde{M}$ denotes the stochastic integral of $\tilde{\gamma}$ with respect to $\{\tilde{M}_t\}$. Also, put

$$\mathcal{E}(\tilde{M})_t := \exp\left(\tilde{M}_t - \frac{1}{2}\langle\tilde{M}\rangle_t\right).$$

Now let $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{P}), (\tilde{\pi}_t, \tilde{I}_t)\}$ be an arbitrary weak solution of the normalized filter equation, and define

$$\tilde{Y}_t^k := \tilde{I}_t^k + \int_0^t \tilde{\pi}_s h^k\, ds, \quad \forall\, t \in [0,T],\ k = 1, \dots, r; \tag{3.2.14}$$

$$\tilde{\chi}_t := \mathcal{E}\left(-\sum_{k=1}^{r}(\tilde{\pi}h^k) \bullet \tilde{I}^k\right)_t, \quad \forall\, t \in [0,T]. \tag{3.2.15}$$

Since $\{\tilde{I}_t,\ t \in [0,T]\}$ is a $\{\tilde{\mathcal{F}}_t\}$-Wiener process, it follows that $\{\tilde{\chi}_t,\ t \in [0,T]\}$ is a continuous strictly-positive $\{\tilde{\mathcal{F}}_t\}$-local martingale on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{P})$, and

$$\frac{1}{\tilde{\chi}_t} = \mathcal{E}\left(\sum_{k=1}^{r}(\tilde{\pi}h^k) \bullet \tilde{Y}^k\right)_t, \quad \forall\, t \in [0,T]. \tag{3.2.16}$$

Define the $\mathcal{M}^+(\mathbb{R}^d)$-valued process $\{\tilde{\sigma}_t,\ t \in [0,T]\}$ on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{P})$ by

$$\tilde{\sigma}_t\phi := \frac{\tilde{\pi}_t\phi}{\tilde{X}_t}, \quad \forall\, t \in [0,T],\ \phi \in B(\mathbb{R}^d). \tag{3.2.17}$$

Hence

$$\tilde{\sigma}_t\phi := (\tilde{\pi}_t\phi)\, \mathcal{E}\left(\sum_{k=1}^{r}(\tilde{\pi}h^k) \bullet \tilde{Y}^k\right)_t, \quad \forall\, t \in [0,T],\ \phi \in B(\mathbb{R}^d), \tag{3.2.18}$$

and, in light of (3.2.12), we see that

$$\tilde{P}\left(\int_0^T \sum_{k=1}^{r}[\tilde{\sigma}_s|h^k\phi + \mathcal{B}_k\phi|]^2\, ds < \infty\right) = 1, \quad \forall\, \phi \in C_c^\infty(\mathbb{R}^d) \cup \{1\}.$$

Using Itô's formula and (3.2.13), we easily verify that the process $\{\tilde{\sigma}_t\}$ given by (3.2.17) satisfies the following stochastic integral relation: for each $\phi \in C_c^\infty(\mathbb{R}^d) \cup \{1\}$ we have

$$\tilde{\sigma}_t\phi = \tilde{\sigma}_0\phi + \int_0^t \tilde{\sigma}_s(\mathcal{A}\phi)\, ds + \sum_{k=1}^{r}\int_0^t \tilde{\sigma}_s(h^k\phi + \mathcal{B}_k\phi)\, d\tilde{Y}_s^k, \quad \forall\, t \in [0,T]. \tag{3.2.19}$$

We will shortly use the form of (3.2.19) to motivate the notion of a weak solution of the unnormalized filter equation.

Remark 3.2.24. From Definition 3.2.19 and (3.2.18) we see that $t \mapsto \tilde{\sigma}_t\phi : [0,T] \to \mathbb{R}$ is continuous for each bounded continuous $\phi : \mathbb{R}^d \to \mathbb{R}$, thus $\{\tilde{\sigma}_t\}$ is a *continuous* $\mathcal{M}^+(\mathbb{R}^d)$-valued process which is $\{\mathcal{F}_t\}$-adapted. Moreover, from (3.2.18), we see that the random element $\tilde{\sigma}_0$ takes values in $\mathcal{P}(\mathbb{R}^d)$, the set of *probability* measures on $\mathbb{R}^d$.

Remark 3.2.25. If, in (3.2.18), we use the optimal filter $\{\pi_t\}$ in place of $\{\tilde{\pi}_t\}$ and the observation process $\{Y_t\}$ in place of $\{\tilde{Y}_t\}$ to get an $\mathcal{M}^+(\mathbb{R}^d)$-valued and $\{\mathcal{F}_t^Y\}$-adapted process $\{\sigma_t\}$, namely

$$\sigma_t\phi := (\pi_t\phi)\, \mathcal{E}\left(\sum_{k=1}^{r}(\pi h^k) \bullet Y^k\right)_t, \quad \forall\, t \in [0,T],\ \phi \in B(\mathbb{R}^d), \tag{3.2.20}$$

then $\{\sigma_t\}$ is the *unnormalized optimal filter* for the filtering problem given by (3.1.2) and (3.1.3) which has been introduced in Section 2.4 (see (2.4.11)).

Remark 3.2.26. In (3.2.19) the "driving process" $\{\tilde{Y}_t\}$ is a continuous $\{\tilde{\mathcal{F}}_t\}$-semimartingale defined on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{P})$ by (3.2.14). The equation (3.2.19) becomes more tractable if we can replace $\tilde{P}$ with some equivalent probability measure $\tilde{Q}$ such that $\{\tilde{Y}_t,\ t \in [0,T]\}$ is an $\{\tilde{\mathcal{F}}_t\}$-Wiener process with respect to $\tilde{Q}$. To this end, observe from (3.2.15) that $\{(\tilde{\chi}_t, \tilde{\mathcal{F}}_t),\ t \in [0,T]\}$ is a continuous local martingale on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$, and that, if it is also a "genuine" martingale, then

$$\tilde{Q}(A) := \mathrm{E}^{\tilde{P}}[\tilde{\chi}_T; A], \quad \forall A \in \tilde{\mathcal{F}}, \tag{3.2.21}$$

defines a probability measure $\tilde{Q}$ on $(\tilde{\Omega}, \tilde{\mathcal{F}})$ which is equivalent to the probability measure $\tilde{P}$, namely

$$\tilde{P} \equiv \tilde{Q} \quad [\tilde{\mathcal{F}}]. \tag{3.2.22}$$

From (3.2.14), (3.2.15), and the Girsanov theorem, it then follows that $\{(\tilde{Y}_t, \tilde{\mathcal{F}}_t),\ t \in [0,T]\}$ is a Wiener process on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{Q})$.

Remark 3.2.27. A sufficient condition on the weak solution $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{P}), (\tilde{\pi}_t, \tilde{I}_t)\}$ and sensor function $h(\cdot)$ which ensures $\{(\tilde{\chi}_t, \tilde{\mathcal{F}}_t),\ t \in [0,T]\}$ is a martingale, is that

$$\mathrm{E}^{\tilde{P}}\left[\exp\left(\frac{1}{2}\sum_{k=1}^{r}\int_0^T [\tilde{\pi}_s h^k]^2\, ds\right)\right] < \infty$$

(see Corollary 3.5.13 of Karatzas and Shreve [16]). In particular, this condition always holds when $h^k \in B(\mathbb{R}^d)$, $1 \leq k \leq r$.

With the preceding discussion in mind, and motivated by (3.2.19), we next formulate the notion of *weak solution of the unnormalized filter equation*, together with pathwise uniqueness and uniqueness in law for this notion of solution (C.f. Définition IV.1 of Szpirglas [36]):

Definition 3.2.28. A pair $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{Q}), (\tilde{\sigma}_t, \tilde{Y}_t)\}$ is a *weak solution* of the unnormalized filter equation when

1. $(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{Q})$ is a complete filtered probability space;

2. $\{\tilde{Y}_t,\ t \in [0,T]\}$ is an $\mathbb{R}^r$-valued $\{\tilde{\mathcal{F}}_t\}$-Wiener process;

3. $\{\tilde{\sigma}_t,\ t \in [0,T]\}$ is a $\mathcal{M}^+(\mathbb{R}^d)$-valued continuous $\{\tilde{\mathcal{F}}_t\}$-adapted process such that the random element $\tilde{\sigma}_0$ takes values in $\mathcal{P}(\mathbb{R}^d)$, and, for each $\phi \in C_c^\infty(\mathbb{R}^d) \cup \{1\}$, we have the following:

   (a)
$$\tilde{Q}\left(\int_0^T \sum_{k=1}^r [\tilde{\sigma}_s|h^k\phi + \mathcal{B}_k\phi|]^2\, ds < \infty\right) = 1; \tag{3.2.23}$$

   (b) the LHS and RHS of (3.2.19) are indistinguishable.

Definition 3.2.29. The unnormalized filter equation has the property of *pathwise uniqueness* when the following holds: If $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{Q}), (\tilde{\sigma}_t^1, \tilde{Y}_t)\}$ and $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{Q}), (\tilde{\sigma}_t^2, \tilde{Y}_t)\}$ are weak solutions of the unnormalized filter equation with $\tilde{Q}(\tilde{\sigma}_0^1 = \tilde{\sigma}_0^2) = 1$, then

$$\tilde{Q}\left(\tilde{\sigma}_t^1 = \tilde{\sigma}_t^2 \quad \forall t \in [0,T]\right) = 1.$$

Definition 3.2.30. The unnormalized filter equation has the property of *uniqueness in joint law* when the following holds: If $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{Q}), (\tilde{\sigma}_t, \tilde{Y}_t)\}$ and $\{(\bar{\Omega}, \bar{\mathcal{F}}, \{\bar{\mathcal{F}}_t\}, \bar{Q}), (\bar{\sigma}_t, \bar{Y}_t)\}$ are weak solutions of the unnormalized filter equation with $\mathcal{L}_{\tilde{Q}}(\tilde{\sigma}_0) = \mathcal{L}_{\bar{Q}}(\bar{\sigma}_0)$, then $\{(\tilde{\sigma}_t, \tilde{Y}_t),\ t \in [0,T]\}$ and $\{(\bar{\sigma}_t, \bar{Y}_t),\ t \in [0,T]\}$ have the same finite-dimensional distributions.

3.3 Main Result

We will now establish that the unnormalized filter equation has the property of pathwise uniqueness in the sense of Definition 3.2.29, and then use this result and a construction of Yamada and Watanabe [37] to see that the normalized filter equation has the property of uniqueness in joint law in the sense of Definition 3.2.22. This is the result we will need to study convergence of nonlinear filters by the method of martingale problems in Chapter 5 (we also automatically get uniqueness in joint law for the unnormalized filter equation in the sense of Definition 3.2.30). To this end we postulate the following conditions on the mappings $b(\cdot)$, $B(\cdot)$, $c(\cdot)$ in (3.1.2), and the mapping $h(\cdot)$ in (3.1.3):

Condition 3.3.31. *The mapping $b : \mathbb{R}^d \to \mathbb{R}^d$ is Borel-measurable, and the mappings $B : \mathbb{R}^d \to \mathbb{R}^{d\times r}$ and $c : \mathbb{R}^d \to \mathbb{R}^{d\times l}$ are continuous. There exists a constant $C \in [0,\infty)$ such that*

$$\max_{i,j,k}\{|b^i(x)|,\ |B^{ij}(x)|, |c^{ik}(x)|\} \leq C[1+|x|], \quad \forall x \in \mathbb{R}^d.$$

Condition 3.3.32. *The mapping $c : \mathbb{R}^d \to \mathbb{R}^{d\times l}$ is such that the matrix $c(x)c^T(x)$ is strictly positive definite for every $x \in \mathbb{R}^d$.*

Condition 3.3.33. *The mapping $h : \mathbb{R}^d \to \mathbb{R}^r$ is Borel-measurable and uniformly bounded.*

We can now state the main result of this chapter:

Theorem 3.3.34. *Suppose that Conditions 3.3.31, 3.3.32, and 3.3.33 hold for the nonlinear filtering problem given by* **E.1**, **E.2** *and* **E.3**. *Then:*

(i) The unnormalized filter equation has the property of pathwise uniqueness;

(ii) The normalized filter equation has the property of uniqueness in joint law;

(iii) The unnormalized filter equation has the property of uniqueness in joint law.

Before establishing Theorem 3.3.34, in the following remarks we first discuss the relation between this uniqueness result and those of Szpirglas [36], as well as other uniqueness results on the equations of nonlinear filtering due to Kurtz and Ocone [22]; Bhatt, Kallianpur and Karandikar [1]; and Rozovskii [33].

Remark 3.3.35. Szpirglas [36] establishes pathwise uniqueness and uniqueness in law for the normalized and unnormalized filter equations corresponding to the following nonlinear filtering problem: The signal $\{X_t\}$ is a homogeneous Markov process with values in a complete separable metric space E with the corresponding Borel semigroup $\{P_t\}$ and weak infinitesimal generator $\mathcal{A}$, and the observation process is

$$Y_t := W_t + \int_0^t h(X_u)\,du, \qquad t \in [0,T],$$

where $\{W_t\}$ is an $\mathbb{R}^{d_1}$-valued Wiener process *independent* of the Markov process $\{X_t\}$, and the sensor function $h : E \to \mathbb{R}^{d_1}$ is uniformly bounded and $\mathcal{B}(E)$-measurable. In this context, by a weak solution of the unnormalized filter equation is meant a pair $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{Q}), (\tilde{\sigma}_t, \tilde{Y}_t)\}$ such that

(a) $(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{Q})$ is a complete filtered probability space;

(b) $\{\tilde{Y}_t,\ t \in [0,T]\}$ is an $\mathbb{R}^{d_1}$-valued $\{\tilde{\mathcal{F}}_t\}$-Wiener process;

(c) $\{\tilde{\sigma}_t,\ t \in [0,T]\}$ is a $\mathcal{M}^+(E)$-valued, corlol, $\{\tilde{\mathcal{F}}_t\}$-adapted process, the random element $\tilde{\sigma}_0$ takes values in $\mathcal{P}(E)$, and $\sup_{t\in[0,T]} \mathrm{E}[|\tilde{\sigma}_t 1|^2] < \infty$;

(d) For each $\phi \in \mathcal{D}(\mathcal{A})$ one has to within indistinguishability that

$$\tilde{\sigma}_t\phi = \tilde{\sigma}_0\phi + \int_0^t \tilde{\sigma}_s(\mathcal{A}\phi)\,ds + \sum_{k=1}^{r} \int_0^t \tilde{\sigma}_s(h^k\phi)\,d\tilde{Y}_s^k, \quad \forall\, t \in [0,T]. \qquad (3.3.24)$$

(See Définition IV.1 of Szpirglas [36]). The nice feature of (3.3.24) is that it includes reference to just one unbounded linear operator, namely the infinitesimal generator $\mathcal{A}$ of the signal process. One can use the resolvent identity of semigroups to eliminate $\mathcal{A}$ and re-write (3.3.24) in the form

$$\tilde{\sigma}_t\phi = \tilde{\sigma}_0(P_t\phi) + \sum_{k=1}^{r}\int_0^t \tilde{\sigma}_s(h^k P_{t-s}\phi)\, d\tilde{Y}_s^k, \quad \forall\, t \in [0,T], \tag{3.3.25}$$

where $\{P_t\}$ is the Borel semigroup with infinitesimal generator $\mathcal{A}$. There is complete equivalence between (3.3.24) and (3.3.25) in the sense that if the pair $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{Q}),$ $(\tilde{\sigma}_t, \tilde{Y}_t)\}$ is subject to (a), (b), (c), then (3.3.24) holds for each $\phi \in \mathcal{D}(\mathcal{A})$ if and only if (3.3.25) holds for each $\phi \in B(E)$ (see Théorème IV.1 of [36]). Consequently, it is enough to establish pathwise uniqueness for (3.3.25) in order to conclude pathwise uniqueness for the unnormalized filter equation. The advantage of (3.3.25) is that it involves only the *bounded* linear operators $\{P_t\}$, and this structure makes it possible to establish pathwise uniqueness for solutions of (3.3.25) by iterating a simple Gronwall-type integral inequality (see Section V.2 of Szpirglas [36]). Comparing (3.3.24) with the unnormalized filter equation (3.2.19) for the nonlinear filtering problem defined by (3.1.2) and (3.1.3), we see that (3.2.19) includes two unbounded linear operators, namely the first-order differential operator $\mathcal{B}_k$ which results from dependence of the signal $\{X_t\}$ on the Wiener process $\{W_t\}$ of the observation equation, as well as the second-order differential operator $\mathcal{A}$ corresponding to the signal process $\{X_t\}$. In this case there seems to be no clear way of adapting the elegant semigroup ideas of [36] to remove both of these unbounded operators and get an equivalent equation involving just bounded linear operators. Furthermore, the requirement that $(\mathcal{D}(\mathcal{A}), \mathcal{A})$ be a weak infinitesimal generator of the semigroup $\{P_t\}$ is rather strong and cannot be secured on the basis of Condition 3.3.31, 3.3.32, and 3.3.33 that we postulate (in fact, on the basis of these conditions we can infer only the *well-posedness of the martingale problem for* $(\mathcal{D}(\mathcal{A}), \mathcal{A})$, which is generally a much weaker property). Accordingly, the approach that we shall use to establish Theorem 3.3.34(i) is very different from that of Szpirglas [36], and relies on a uniqueness theorem for measure-valued evolution equations (see Theorem 3.4.44 to follow).

Remark 3.3.36. Uniqueness for the normalized and unnormalized filter equations has also been studied by Bhatt, Kallianpur and Karandikar [1], Kurtz and Ocone [22], and Rozovskii [33] from a somewhat different point of view than that taken by Szpirglas [36] and the present work. To see this in the context of the filtering problem given by (3.1.2) and (3.1.3), observe from Remark 3.2.25 that the unnormalized optimal filter $\{\sigma_t\}$ solves the Duncan-Mortensen-Zakai equation, namely for each $\phi \in C_c^\infty(\mathbb{R}^d) \cup \{1\}$ we have

$$\sigma_t\phi = \pi_0\phi + \int_0^t \sigma_s(\mathcal{A}\phi)\, ds + \sum_{k=1}^{d_1} \int_0^t \sigma_s(h^k\phi + \mathcal{B}_k\phi)\, dY_s^k, \quad \forall\, t \in [0,T]. \quad (3.3.26)$$

With this in mind, the following question is natural: Suppose that $\{\rho_t\}$ is some $\mathcal{M}^+(\mathbb{R}^d)$-valued, corlol, and $\{\mathcal{F}_t^Y\}$*-adapted* process on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$, such that for each $\phi \in C_c^\infty(\mathbb{R}^d) \cup \{1\}$ we have

$$\rho_t\phi = \pi_0\phi + \int_0^t \rho_s(\mathcal{A}\phi)\, ds + \sum_{k=1}^{d_1} \int_0^t \rho_s(h^k\phi + \mathcal{B}_k\phi)\, dY_s^k, \quad \forall\, t \in [0,T]. \quad (3.3.27)$$

Does it follow that $\{\sigma_t\}$ and $\{\rho_t\}$ are indistinguishable? The works of Bhatt, Kallianpur and Karandikar ([1], Theorem 3.1), Kurtz and Ocone ([22], Theorems 4.2 and 4.7), and Rozovskii ([33], Theorem 3.1) provide conditions on the nonlinear filtering problem for which the answer is in the affirmative. Uniqueness in this sense is useful for the following reason: the observation process $\{Y_t\}$ is the random data that "drives" the unnormalized filter equation (3.3.26), and if we can "non-anticipatively" use the individual paths of $\{Y_t\}$ as data to compute a measure-valued process $\{\rho_t\}$ which satisfies (3.3.27) — e.g. by a numerical method — then uniqueness ensures that $\{\rho_t\}$ is in fact the desired unnormalized optimal filter $\{\sigma_t\}$. It should be noted that uniqueness in this sense can be established for much more general nonlinear filtering problems than that represented by the simple model (3.1.2) and (3.1.3). In fact, Theorem 3.1 of [1] deals with a filtering problem in which the signal process takes values in a complete separable metric space (not necessarily locally compact), the sensor function $h(\cdot)$ need not be uniformly bounded but only satisfy a finite-energy condition similar to (3.2.4), the dependence of the signal $\{X_t\}$ on the Wiener process $\{W_t\}$ is more general than that given by the explicit model (3.1.2), (3.1.3) (see (1.3) of [1]), and the joint signal/observation process $\{(X_t, Y_t)\}$ is the corlol solution of a well-posed martingale problem.

The sense of pathwise uniqueness in the preceding paragraph is different from that established by Theorem 3.3.34(i), since the candidate solution $\{\rho_t\}$ of the filter equation (3.3.27) is postulated to be adapted specifically to the observation filtration $\{\mathcal{F}_t^Y\}$ (in fact, the arguments used in [1], [22], and [33] rely crucially on this restriction). In contrast, Theorem 3.3.34(i) establishes pathwise uniqueness in the more general sense of Definition 3.2.29, where the candidate solutions $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{Q}), (\tilde{\sigma}_t^1, \tilde{Y}_t)\}$ and $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{Q}), (\tilde{\sigma}_t^2, \tilde{Y}_t)\}$ are defined on an arbitrary filtered probability space, and there is no insistence that the measure-valued components $\{\tilde{\sigma}_t^1\}$ and $\{\tilde{\sigma}_t^2\}$ of the two solutions be adapted to the self-filtration of $\{\tilde{Y}_t\}$. The usefulness of this latter notion of pathwise uniqueness is that, by an adaptation to the filter equations of the construction of Yamada and Watanabe [37], it leads to uniqueness in law for the normalized and unnormalized filter equations (see Theorem 3.3.34(ii) and (iii), and recall Definitions 3.2.22 and 3.2.30). Uniqueness in law turns out to be essential for studying weak limits and approximations of the nonlinear filter equations by the method of martingale problems and weak convergence.

Remark 3.3.37. Using the method of stochastic flows, Kunita ([18], Theorem 6.2.8) establishes a form of pathwise uniqueness for the unnormalized filter equation. Rather restrictive boundedness and smoothness conditions on the coefficients of (3.1.2) and (3.1.3) appear necessary for this method to work.

A basic property of Itô stochastic differential equations due to Yamada and Watanabe [37] is that pathwise uniqueness implies uniqueness in joint law, so that pathwise uniqueness is the stronger of the two uniqueness properties. It turns out that the basic Yamada-Watanabe argument extends to the measure-valued filter equations, so that pathwise uniqueness is again the stronger property (this is how we will conclude (ii) and (iii) from (i) in Theorem 3.3.34). Linearity of the unnormalized filter equation in fact implies the converse, so that for this particular equation the two uniqueness properties are actually equivalent:

Theorem 3.3.38. *Suppose that Conditions 3.3.31, 3.3.32, and 3.3.33 hold for the nonlinear filtering problem given by* **E.1**, **E.2** *and* **E.3**. *Then uniqueness in joint law implies pathwise uniqueness for the unnormalized filter equation.*

3.4 Proofs of Theorems 3.3.34 and 3.3.38:

The terminology in the next remark will be useful for dealing with measure-valued evolution equations:

Remark 3.4.39. Suppose that E is a complete separable metric space, and $\mathcal{Q} : \mathcal{D}(\mathcal{Q}) \to B(E)$ is a mapping with domain $\mathcal{D}(\mathcal{Q}) \subset B(E)$. Then $\{\mu_t,\ t \in [0,\infty)\}$ is an $\mathcal{M}^+(E)$-valued solution of the evolution equation for $(\mathcal{Q},\mathcal{D}(\mathcal{Q}))$ when (i) $\mu_t \in \mathcal{M}^+(E)$, $\forall\ t \in [0,\infty)$, and $\mu_0 \in \mathcal{P}(E)$; (ii) for each $\Gamma \in \mathcal{B}(E)$, the mapping $t \mapsto \mu_t(\Gamma) : [0,\infty) \to [0,\infty)$ is Borel-measurable; (iii) for each $f \in \mathcal{D}(\mathcal{Q})$ we have $\int_0^t |\mu_s(\mathcal{Q}f)|ds < \infty$, $\forall t \in [0,\infty)$, and

$$\mu_t f = \mu_0 f + \int_0^t \mu_s(\mathcal{Q}f)\,ds, \quad \forall t \in [0,\infty). \tag{3.4.28}$$

Moreover, $\{\mu_t,\ t \in [0,\infty)\}$ is called a $\mathcal{P}(E)$-valued solution of the evolution equation for $(\mathcal{Q},\mathcal{D}(\mathcal{Q}))$ when it is an $\mathcal{M}^+(E)$-valued solution with $\mu_t(E) = 1$,

$\forall\ t \in [0, \infty)$. The evolution equation for $(\mathcal{Q}, \mathcal{D}(\mathcal{Q}))$ is said to have uniqueness in the class of $\mathcal{M}^+(E)$-valued solutions over the interval $[0, \infty)$ when, for any two such solutions $\{\mu_t^i,\ t \in [0, \infty)\}$, $i = 1, 2$, with $\mu_0^1 = \mu_0^2$, it follows that $\mu_t^1 = \mu_t^2$, $\forall t \in [0, \infty)$. The notion of uniqueness within the class of $\mathcal{P}(E)$-valued solutions over the interval $[0, \infty)$ has an analogous formulation. Finally, the preceding terminology adapts in an obvious way to solutions $\{\mu_t,\ t \in [0, T]\}$ defined over the finite interval $[0, T]$.

Proof of Theorem 3.3.34(i) We shall need the following result, the proof of which is given in Appendix A.1:

Fact 3.4.40. *Suppose that Conditions 3.3.31-3.3.33 hold and let* $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{Q}), (\tilde{\sigma}, \tilde{Y})\}$ *be a weak solution of the unnormalized filter equation. Then, for every* $\alpha \in (1, \infty)$ *there exists a constant* $\gamma(\alpha) \in [0, \infty)$ *such that*

$$\mathrm{E}^{\tilde{Q}}\left[\sup_{0 \le s \le T} |\tilde{\sigma}_s 1|^\alpha\right] \le \gamma(\alpha). \tag{3.4.29}$$

Now fix two weak solutions $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{Q}), (\tilde{\sigma}_t^i, \tilde{Y}_t)\}$, $i = 1, 2$, of the unnormalized filter equation, such that

$$\tilde{Q}\left[\tilde{\sigma}_0^1 = \tilde{\sigma}_0^2\right] = 1, \tag{3.4.30}$$

and define product measures on $(\mathbb{R}^{2d}, \mathcal{B}(\mathbb{R}^{2d}))$ by

$$\mu_t^{12}(\cdot, \tilde{\omega}) := (\tilde{\sigma}_t^1 \times \tilde{\sigma}_t^2)(\cdot, \tilde{\omega}), \quad \forall (t, \tilde{\omega}) \in [0, T] \times \tilde{\Omega}.$$

A simple application of the Dynkin class theorem establishes

Fact 3.4.41. *For every* $\Gamma \in \mathcal{B}(\mathbb{R}^{2d})$, *the mapping* $(t, \tilde{\omega}) \mapsto \mu_t^{12}(\Gamma, \tilde{\omega}) : \tilde{\Omega} \times [0, T] \to [0, \infty)$ *is measurable with respect to the* $\{\tilde{\mathcal{F}}_t\}$*-progressive* σ*-algebra.*

Also put

$$\nu_t^{12}(\Gamma) := \mathrm{E}^{\tilde{Q}}[\mu_t^{12}(\Gamma)], \quad \forall \Gamma \in \mathcal{B}(\mathbb{R}^{2d}),\ t \in [0, T]. \tag{3.4.31}$$

It readily follows that ν_t^{12} defines a (positive) measure on $(\mathbb{R}^{2d}, \mathcal{B}(\mathbb{R}^{2d}))$ for every $t \in [0,T]$. By Fact 3.4.40,

$$\begin{aligned}\nu_t^{12}(\mathbb{R}^{2d}) &= \mathrm{E}^{\tilde{Q}}[(\tilde{\sigma}_t^1 1)(\tilde{\sigma}_t^2 1)] \\ &\leq \left(\mathrm{E}^{\tilde{Q}}[\sup_{0\leq s\leq T} |\tilde{\sigma}_s^1 1|^2]\right)^{\frac{1}{2}} \left(\mathrm{E}^{\tilde{Q}}[\sup_{0\leq s\leq T} |\tilde{\sigma}_s^2 1|^2]\right)^{\frac{1}{2}} \leq \gamma(2), \quad \forall t \in [0,T].\end{aligned} \tag{3.4.32}$$

This shows that ν_t^{12} is a positive measure on $(\mathbb{R}^{2d}, \mathcal{B}(\mathbb{R}^{2d}))$, uniformly bounded with respect to $t \in [0,T]$, while Fact 3.4.41 with Fubini's theorem shows that the mapping $t \mapsto \nu_t^{12}(\Gamma) : [0,T] \to \mathbb{R}$ is Borel-measurable for each $\Gamma \in \mathcal{B}(\mathbb{R}^{2d})$. Next, define $\nu_t^{11}, \nu_t^{22} \in \mathcal{M}^+(\mathbb{R}^{2d})$, $t \in [0,T]$, analogously to ν_t^{12}, by

$$\nu_t^{ii}(\Gamma) := \mathrm{E}^{\tilde{Q}}[(\tilde{\sigma}_t^i \times \tilde{\sigma}_t^i)(\Gamma)], \quad \forall \Gamma \in \mathcal{B}(\mathbb{R}^{2d}),\ t \in [0,T],\ i = 1,2. \tag{3.4.33}$$

In the same way as for $\nu_\cdot^{12}$, we see that $\nu_\cdot^{ii}$ are positive measures on $(\mathbb{R}^{2d}, \mathcal{B}(\mathbb{R}^{2d}))$, uniformly bounded with respect to $t \in [0,T]$, and the mappings $t \mapsto \nu_t^{ii}(\Gamma) : [0,T] \to \mathbb{R}$ are Borel-measurable for each $\Gamma \in \mathcal{B}(\mathbb{R}^{2d})$, $i = 1,2$.

Remark 3.4.42. For mappings $f_1, f_2 \in B(\mathbb{R}^d)$ define the tensor product of f_1 with f_2 to be the mapping $f_1 \otimes f_2 : \mathbb{R}^{2d} \to \mathbb{R}$ given by

$$f_1 \otimes f_2(x_1, x_2) := f_1(x_1) f_2(x_2), \quad \forall\, x_1, x_2 \in \mathbb{R}^d.$$

In view of (3.4.31) and (3.4.33), for each $f_1, f_2 \in B(\mathbb{R}^d)$ we have

$$\nu_t^{12}(f_1 \otimes f_2) = \mathrm{E}^{\tilde{Q}}[(\tilde{\sigma}_t^1 f_1)(\tilde{\sigma}_t^2 f_2)], \tag{3.4.34}$$

$$\nu_t^{ii}(f_1 \otimes f_2) = \mathrm{E}^{\tilde{Q}}[(\tilde{\sigma}_t^i f_1)(\tilde{\sigma}_t^i f_2)], \quad i = 1,2. \tag{3.4.35}$$

From (3.4.30), (3.4.31), and (3.4.33) we see that

$$\nu_0^{11}, \nu_0^{22} \text{ and } \nu_0^{12} \text{ are probability measures on } \mathcal{B}(\mathbb{R}^{2d}) \text{ and } \nu_0^{11} = \nu_0^{22} = \nu_0^{12}. \tag{3.4.36}$$

Using this fact, we shall establish

$$\nu_t^{11} = \nu_t^{22} = \nu_t^{12}, \quad \forall t \in [0,T], \tag{3.4.37}$$

from which pathwise uniqueness follows. Indeed, if (3.4.37) holds, then for each $f \in B(\mathbb{R}^d)$ we have

$$\nu_t^{11}(f \otimes f) = \nu_t^{22}(f \otimes f) = \nu_t^{12}(f \otimes f), \quad \forall t \in [0,T],$$

and therefore from (3.4.34) and (3.4.35),

$$\begin{aligned} \mathrm{E}^{\tilde{Q}}[(\tilde{\sigma}_t^1 f - \tilde{\sigma}_t^2 f)^2] &= \mathrm{E}^{\tilde{Q}}[(\tilde{\sigma}_t^1 f)(\tilde{\sigma}_t^1 f)] - 2\,\mathrm{E}^{\tilde{Q}}[(\tilde{\sigma}_t^1 f)(\tilde{\sigma}_t^2 f)] + \mathrm{E}^{\tilde{Q}}[(\tilde{\sigma}_t^2 f)(\tilde{\sigma}_t^2 f)] \\ &= \nu_t^{11}(f \otimes f) - 2\nu_t^{12}(f \otimes f) + \nu_t^{22}(f \otimes f) = 0. \end{aligned}$$

Thus, for each $t \in [0,T]$ and $f \in B(\mathbb{R}^d)$, we have

$$\tilde{Q}\left[\tilde{\sigma}_t^1 f = \tilde{\sigma}_t^2 f\right] = 1. \tag{3.4.38}$$

Now $\hat{C}(\mathbb{R}^d)$ equipped with the supremum norm $\|\cdot\|$ is separable. Thus, from (3.4.38), for each $t \in [0,T]$ there is a $\tilde{Q}$-null event $N_t \in \tilde{\mathcal{F}}$ such that, for each $\tilde{\omega} \notin N_t$, we have

$$\tilde{\sigma}_t^1(\tilde{\omega}) f = \tilde{\sigma}_t^2(\tilde{\omega}) f, \quad \forall\, f \in \hat{C}(\mathbb{R}^d). \tag{3.4.39}$$

But $\hat{C}(\mathbb{R}^d)$ is separating for $\mathcal{M}^+(\mathbb{R}^d)$ (see Definition D.1.123 and Fact D.1.129(ii)), thus (3.4.39) establishes $\tilde{Q}[\tilde{\sigma}_t^1 = \tilde{\sigma}_t^2] = 1$ for each $t \in [0,T]$. Now Theorem 3.3.34(i) follows from the fact that $\{\tilde{\sigma}_t^i,\ t \in [0,T]\}$ are continuous (recall Definition 3.2.28).

It therefore remains to establish (3.4.37) in order to prove Theorem 3.3.34(i). To this end, for each $x_1, x_2 \in \mathbb{R}^d$ define the $2d \times 2d$ matrix $\bar{a}(x_1, x_2)$, the $2d$ vector

$\bar{b}(x_1, x_2)$, and the real number $\bar{h}(x_1, x_2)$ by

$$\bar{a}(x_1, x_2) := \begin{bmatrix} cc^T(x_1) & 0 \\ 0 & cc^T(x_2) \end{bmatrix} + \begin{bmatrix} B(x_1) \\ B(x_2) \end{bmatrix} \begin{bmatrix} B^T(x_1) & B^T(x_2) \end{bmatrix} \tag{3.4.40a}$$

$$\bar{b}(x_1, x_2) := \begin{bmatrix} b(x_1) + B(x_1)h(x_2) \\ b(x_2) + B(x_2)h(x_1) \end{bmatrix} \tag{3.4.40b}$$

$$\bar{h}(x_1, x_2) := \sum_{k=1}^{r} h^k(x_1)h^k(x_2). \tag{3.4.40c}$$

Observe that the matrix $\bar{a}(x_1, x_2)$ is symmetric and strictly positive-definite (see Condition 3.3.32), and let $\bar{A}$ be the second order linear differential operator corresponding to the matrices $\bar{a}$ and $\bar{b}$, namely

$$\bar{A}\phi(x) := \sum_{i=1}^{2d} \bar{b}^i(x)\partial_i\phi(x) + \frac{1}{2}\sum_{i,j=1}^{2d} \bar{a}^{ij}(x)\partial_i\partial_j\phi(x), \quad \forall\, \phi \in C^\infty(\mathbb{R}^{2d}),\ x \in \mathbb{R}^{2d}. \tag{3.4.41}$$

From (3.4.40a), (3.4.40b), Condition 3.3.31, and Condition 3.3.33, there is a constant $K \in [0, \infty)$ such that

$$\max_i |\bar{b}^i(x)| \le K[1 + |x|], \quad \max_{i,j} |\bar{a}^{ij}(x)| \le K[1 + |x|^2], \quad \forall x \in \mathbb{R}^{2d}, \tag{3.4.42}$$

and the operator $\bar{A}$ has the following property, which is established in Section 3.5:

Lemma 3.4.43. *Suppose that Conditions 3.3.31–3.3.33 hold. Then* $\{\nu_t^{11},\ t \in [0, T]\}$, $\{\nu_t^{12},\ t \in [0, T]\}$, *and* $\{\nu_t^{22},\ t \in [0, T]\}$, *given by* (3.4.31) *and* (3.4.33), *are* $\mathcal{M}^+(\mathbb{R}^{2d})$*-valued solutions of the evolution equation for* $(\bar{A} + \bar{h}, \mathrm{span}\{1, C_c^\infty(\mathbb{R}^{2d})\})$.

It remains to show that the evolution equation for $(\bar{A} + \bar{h}, \mathrm{span}\{1, C_c^\infty(\mathbb{R}^{2d})\})$ has uniqueness in the class of $\mathcal{M}^+(\mathbb{R}^{2d})$-valued solutions over the interval $[0, T]$, since this fact, along with (3.4.36) and Lemma 3.4.43, gives (3.4.37), as required to establish Theorem 3.3.34(i). To this end we need the following result on uniqueness of measure-valued solutions of the evolution equation corresponding to a multiplicatively perturbed linear second-order differential operator on Euclidean space:

Theorem 3.4.44. *Let $\mathcal{C}$ be the linear second-order differential operator on the finite-dimensional Euclidean space $\mathbb{R}^q$ defined by*

$$\mathcal{D}(\mathcal{C}) := \operatorname{span}\{1, C_c^\infty(\mathbb{R}^q)\}; \tag{3.4.43a}$$

$$\mathcal{C}f(x) := \sum_i \beta^i(x)\partial_i f(x) + \frac{1}{2}\sum_{i,j} \alpha^{ij}(x)\partial_i\partial_j f(x), \quad \forall\, x \in \mathbb{R}^q, \forall\, f \in \mathcal{D}(\mathcal{C}), \tag{3.4.43b}$$

where $\beta : \mathbb{R}^q \to \mathbb{R}^q$ is Borel measurable, $\alpha : \mathbb{R}^q \to \mathbb{S}_{++}^{q\times q}$ (the set of symmetric strictly positive definite $q \times q$ matrices) is continuous, and there exists a constant $K \in [0, \infty)$ such that

$$|\beta^i(x)| \leq K(1 + |x|), \quad |\alpha^{ij}(x)| \leq K(1 + |x|^2), \quad \forall x \in \mathbb{R}^q. \tag{3.4.44}$$

If $\lambda \in B(\mathbb{R}^q)$ then the evolution equation for the operator $(\mathcal{C} - \lambda, \mathcal{D}(\mathcal{C}))$ has uniqueness in the class of $\mathcal{M}^+(\mathbb{R}^q)$-valued solutions over the interval $[0, T]$.

To complete the proof of Theorem 3.3.34(i) we note from (3.4.40) that $\bar{a}(\cdot)$ is continuous on $\mathbb{R}^{2d}$, $\bar{b}(\cdot)$ is Borel-measurable on $\mathbb{R}^{2d}$, and $\bar{h} \in B(\mathbb{R}^{2d})$. That the evolution equation for the operator $(\bar{\mathcal{A}}+\bar{h}, \operatorname{span}\{1, C_c^\infty(\mathbb{R}^{2d})\})$ has uniqueness in the class of $\mathcal{M}^+(\mathbb{R}^{2d})$-valued solutions over the interval $[0, T]$ now follows from (3.4.42) and Theorem 3.4.44 with $q := 2d$, $\beta(\cdot) := \bar{b}(\cdot)$, $\alpha(\cdot) := \bar{a}(\cdot)$, and $\lambda(\cdot) := -\bar{h}(\cdot)$. □

Remark 3.4.45. When $\beta(\cdot)$ and $\lambda(\cdot)$ in Theorem 3.4.44 are *continuous* then $\mathcal{C}$ is a linear operator on $\bar{C}(\mathbb{R}^q)$, and Theorem 3.4.44 is just a very special consequence of a general theorem of Bhatt and Karandikar (see Theorem 3.4 and Remark 1 of [4]) on uniqueness of measure-valued solutions of perturbed evolution equations. However, when $\beta(\cdot)$ and $\lambda(\cdot)$ are only Borel-measurable, then $\mathcal{C}f(\cdot)$ is not continuous for $f \in \mathcal{D}(\mathcal{C})$, and we cannot directly use the result of [4]. We shall use a result of Kurtz and Stockbridge (see [21], Theorem 2.7(c)) on uniqueness of forward equations to establish Theorem 3.4.44 in Appendix A.2.

Remark 3.4.46. The proof just given for Theorem 3.3.34(i) relies on the special structure of the unnormalized filter equation (3.2.19) and does not appear to extend to the normalized filter equation (3.2.11). We have therefore not been able to establish pathwise uniqueness in the sense of Definition 3.2.21 under conditions comparable to those of Theorem 3.3.34.

Proof of Theorem 3.3.34(ii): Let $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{P}), (\tilde{\pi}_t, \tilde{I}_t)\}$ and $\{(\bar{\Omega}, \bar{\mathcal{F}}, \{\bar{\mathcal{F}}_t\}, \bar{P}), (\bar{\pi}_t, \bar{I}_t)\}$ be two weak solutions of the normalized filter equation. By an argument similar to that used for Proposition IX.1.4 of Revuz and Yor [30], to establish uniqueness in joint law it is enough to show that the processes $\{(\tilde{\pi}_t, \tilde{I}_t)\}$ and $\{(\bar{\pi}_t, \bar{I}_t)\}$ are identically distributed when $\tilde{\pi}_0$ and $\bar{\pi}$ are non-random, namely

$$\tilde{\pi}_0 = \bar{\pi}_0 = \mu, \quad \text{for arbitrary } \mu \in \mathcal{P}(\mathbb{R}^d). \tag{3.4.45}$$

Thus suppose (3.4.45) holds for some $\mu \in \mathcal{P}(\mathbb{R}^d)$. Put

$$\tilde{\chi}_t := \mathcal{E}\left(-\sum_{k=1}^{r}(\tilde{\pi} h^k) \bullet \tilde{I}^k\right)_t \qquad \text{and} \qquad \bar{\chi}_t := \mathcal{E}\left(-\sum_{k=1}^{r}(\bar{\pi} h^k) \bullet \bar{I}^k\right)_t, \quad \forall t \in [0,T],$$

and define the measures $\tilde{Q}$ and $\bar{Q}$ on the measurable spaces $(\tilde{\Omega}, \tilde{\mathcal{F}})$ and $(\bar{\Omega}, \bar{\mathcal{F}})$ respectively by

$$\tilde{Q}(A) := \mathrm{E}^{\tilde{P}}[\tilde{\chi}_T; A], \quad \forall A \in \tilde{\mathcal{F}}, \tag{3.4.46}$$

$$\bar{Q}(A) := \mathrm{E}^{\bar{P}}[\bar{\chi}_T; A], \quad \forall A \in \bar{\mathcal{F}}. \tag{3.4.47}$$

Then, with

$$\tilde{Y}_t := \tilde{I}_t + \sum_{k=1}^{r} \int_0^t \tilde{\pi}_u h^k \, du, \qquad \bar{Y}_t := \bar{I}_t + \sum_{k=1}^{r} \int_0^t \bar{\pi}_u h^k \, du, \quad t \in [0,T], \tag{3.4.48}$$

and

$$\tilde{\sigma}_t := \tilde{\pi}_t / \tilde{\chi}_t, \qquad \bar{\sigma}_t := \bar{\pi}_t / \bar{\chi}_t, \quad \forall t \in [0,T], \tag{3.4.49}$$

we see, as in Remark 3.2.23 and Remark 3.2.26, that the pairs $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{Q}), (\tilde{\sigma}_t, \tilde{Y}_t)\}$ and $\{(\bar{\Omega}, \bar{\mathcal{F}}, \{\bar{\mathcal{F}}_t\}, \bar{Q}), (\bar{\sigma}_t, \bar{Y}_t)\}$ are weak solutions of the unnormalized filter equation, with

$$\tilde{\sigma}_0 = \bar{\sigma}_0 = \mu.$$

Define

$$\hat{\Omega} := C_{\mathcal{M}^+(\mathbb{R}^d)}[0,T] \times C_{\mathcal{M}^+(\mathbb{R}^d)}[0,T] \times C_{\mathbb{R}^r}[0,T],$$

which is a complete separable metric space with the usual product metric, and let $\hat{\omega} = (\omega^1, \omega^2, \omega^3)$ be a generic member of $\hat{\Omega}$. By the Yamada-Watanabe construction (see Theorem IV.1.1 of Ikeda and Watanabe [13]), there exists $\hat{P} \in \mathcal{P}(\hat{\Omega})$ such that

YW.1: $\mathcal{L}_{\hat{P}}(\omega^1, \omega^3) = \mathcal{L}_{\tilde{Q}}(\tilde{\sigma}, \tilde{Y})$;

YW.2: $\mathcal{L}_{\hat{P}}(\omega^2, \omega^3) = \mathcal{L}_{\bar{Q}}(\bar{\sigma}, \bar{Y})$;

YW.3: If $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$ is the completion of $(\hat{\Omega}, \mathcal{B}(\hat{\Omega}), \hat{P})$, and $\hat{\mathcal{F}}_t$ is the augmentation of the σ-algebra $\sigma\{\hat{\omega}(s),\ s \in [0,t]\}$ with the null events of $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$, $\forall\, t \in [0,T]$, then $\{\omega_t^3,\ t \in [0,T]\}$ is an $\{\hat{\mathcal{F}}_t\}$-Wiener process on $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$.

From (YW.1), (YW.2), and (YW.3), along with Exercise IV.5.16 of Revuz and Yor [30], it follows that $\{(\hat{\Omega}, \hat{\mathcal{F}}, \{\hat{\mathcal{F}}_t\}, \hat{P}), (\omega^1, \omega^3)\}$ and $\{(\hat{\Omega}, \hat{\mathcal{F}}, \{\hat{\mathcal{F}}_t\}, \hat{P}), (\omega^2, \omega^3)\}$ are weak solutions for the unnormalized filter equation with

$$\omega_0^1 = \omega_0^2 = \mu,$$

and hence, from Theorem 3.3.34(i),

$$\hat{P}(\omega_t^1 = \omega_t^2 \quad \forall t \in [0,T]) = 1. \tag{3.4.50}$$

From (3.4.49) we see that

$$\tilde{\pi}_t\phi = (\tilde{\sigma}_t\phi)/(\tilde{\sigma}_t 1), \quad \forall\, t \in [0,T], \quad \phi \in B(\mathbb{R}^d), \tag{3.4.51}$$

and so, from (3.4.48),

$$\tilde{I}_t^k = \tilde{Y}_t^k - \sum_{k=1}^{r} \int_0^t (\tilde{\sigma}_u h^k)/(\tilde{\sigma}_u 1)\, du, \quad \forall\, k = 1, 2, \ldots, r, \quad t \in [0, T]. \tag{3.4.52}$$

From (3.4.51) and (3.4.52) there exists a measurable mapping $\Phi : C_{\mathcal{M}^+(\mathbb{R}^d)}[0,T] \times C_{\mathbb{R}^r}[0,T] \to C_{\mathcal{P}(\mathbb{R}^d)}[0,T] \times C_{\mathbb{R}^r}[0,T]$ such that

$$(\tilde{\pi}, \tilde{I}) = \Phi(\tilde{\sigma}, \tilde{Y}). \tag{3.4.53}$$

Now (3.4.51) and (3.4.52) continue to hold with "overbar" in place of "tilde", and hence

$$(\bar{\pi}, \bar{I}) = \Phi(\bar{\sigma}, \bar{Y}). \tag{3.4.54}$$

Thus, for each $\Gamma \in \mathcal{B}(C_{\mathcal{P}(\mathbb{R}^d)}[0,T] \times C_{\mathbb{R}^r}[0,T])$, we see from (3.4.46), (3.4.53), and (YW.1), that

$$\tilde{P}((\tilde{\pi}, \tilde{I}) \in \Gamma) = \mathrm{E}^{\tilde{Q}}[(\tilde{\sigma}_T 1)^{-1} I_\Gamma(\Phi(\tilde{\sigma}, \tilde{Y}))] = \mathrm{E}^{\hat{P}}[(\omega_T^1 1)^{-1} I_\Gamma(\Phi(\omega^1, \omega^3))], \tag{3.4.55}$$

and, from (3.4.47), (3.4.54), and (YW.2), we similarly have

$$\bar{P}((\bar{\pi}, \bar{I}) \in \Gamma) = \mathrm{E}^{\bar{Q}}[(\bar{\sigma}_T 1)^{-1} I_\Gamma(\Phi(\bar{\sigma}, \bar{Y}))] = \mathrm{E}^{\hat{P}}[(\omega_T^2 1)^{-1} I_\Gamma(\Phi(\omega^2, \omega^3))]. \tag{3.4.56}$$

Now (3.4.50), (3.4.55), and (3.4.56) show that $\tilde{P}((\tilde{\pi}, \tilde{I}) \in \Gamma) = \bar{P}((\bar{\pi}, \bar{I}) \in \Gamma)$, as required. □

Proof of Theorem 3.3.34(iii): The proof is an obvious simplification of the proof of Theorem 3.3.34(ii) and is omitted. □

Proof of Theorem 3.3.38: Let $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{Q}), (\tilde{\sigma}_t^i, \tilde{Y}_t)\}$, $i = 1, 2$ be two weak solutions of the unnormalized filter equation. Define

$$\tilde{\sigma}_t^3(\cdot) := \frac{\tilde{\sigma}_t^1(\cdot) + \tilde{\sigma}_t^2(\cdot)}{2}, \quad t \in [0, T].$$

It then follows that $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{Q}), (\tilde{\sigma}_t^3, \tilde{Y}_t)\}$ is a weak solution of the unnormalized filter equation. Therefore, the postulated uniqueness in joint law together with Fact 3.4.40 implies that for an arbitrary $\phi \in B(\mathbb{R}^d)$ we have

$$2\,\mathrm{E}^{\tilde{Q}}\left[\left(\frac{\tilde{\sigma}_t^1\phi + \tilde{\sigma}_t^2\phi}{2}\right)^2\right] - \mathrm{E}^{\tilde{Q}}[(\tilde{\sigma}_t^1\phi)^2] - \mathrm{E}^{\tilde{Q}}[(\tilde{\sigma}_t^2\phi)^2] = 0, \quad \forall t \in [0,T],$$

hence, rearranging, gives

$$\mathrm{E}^{\tilde{Q}}[(\tilde{\sigma}_t^1\phi - \tilde{\sigma}_t^2\phi)^2] = 0, \quad \forall t \in [0,T], \quad \phi \in \hat{C}(\mathbb{R}^d), \tag{3.4.57}$$

and therefore

$$\tilde{Q}[\tilde{\sigma}_t^1\phi = \tilde{\sigma}_t^2\phi] = 1, \quad \forall t \in [0,T], \quad \phi \in B(\mathbb{R}^d).$$

Since $\hat{C}(\mathbb{R}^d)$ is separating for $\mathcal{M}^+(\mathbb{R}^d)$ (see Definition D.1.123 and Fact D.1.129(ii)), we have $\tilde{Q}[\tilde{\sigma}_t^1 = \tilde{\sigma}_t^2] = 1$ for each $t \in [0,T]$. Now Theorem 3.3.34(i) follows from the fact that $\{\tilde{\sigma}_t^i,\ t \in [0,T]\}$ are continuous (recall Definition 3.2.28). $\square$

3.5 Proof of Lemma 3.4.43

For arbitrary $f_1, f_2 \in C_c^\infty(\mathbb{R}^d)$ put

$$\begin{aligned}\tilde{\mathcal{A}}(f_1 \otimes f_2) := f_1 \otimes (\mathcal{A}f_2) + (\mathcal{A}f_1) \otimes f_2 + \sum_{k=1}^{r}\big[(h^k f_1) \otimes (h^k f_2) + (h^k f_1) \otimes (\mathcal{B}_k f_2) \\ + (\mathcal{B}_k f_1) \otimes (h^k f_2) + (\mathcal{B}_k f_1) \otimes (\mathcal{B}_k f_2)\big],\end{aligned} \tag{3.5.58}$$

where $\mathcal{A}$ and $\mathcal{B}_k$ are given by (3.2.8). We need the following lemmas, the proofs of which are given in Appendix A.1:

Lemma 3.5.47. *Suppose Conditions 3.3.31–3.3.33 hold, let $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{Q}), (\tilde{\sigma}_t^i, \tilde{Y}_t)\}$, $i = 1, 2$, be weak solutions of the unnormalized filter equation, and define the*

$\mathcal{M}^+(\mathbb{R}^{2d})$-valued functions $\{\nu_t^{12},\ t \in [0,T]\}$, $\{\nu_t^{11},\ t \in [0,T]\}$, and $\{\nu_t^{22},\ t \in [0,T]\}$, as in (3.4.31) *and* (3.4.33). *Then, for each $f_1, f_2 \in C_c^\infty(\mathbb{R}^d)$, we have*

$$\nu_t^{12}(f_1 \otimes f_2) = \nu_0^{12}(f_1 \otimes f_2) + \int_0^t \nu_u^{12}(\tilde{A}(f_1 \otimes f_2))\, du, \quad \forall\, t \in [0,T], \tag{3.5.59}$$

with identical relations for $\nu_\cdot^{11}$ and $\nu_\cdot^{22}$ in place of $\nu_\cdot^{12}$.

Lemma 3.5.48. *For $\tilde{A}$ and $\bar{A}$ defined in* (3.5.58) *and* (3.4.41) *respectively, we have*

$$\tilde{A}(f_1 \otimes f_2)(x) = \bar{A}(f_1 \otimes f_2)(x) + \bar{h}(x)(f_1 \otimes f_2)(x), \quad \forall\, x \in \mathbb{R}^{2d}, \tag{3.5.60}$$

for each $f_1, f_2 \in C_c^\infty(\mathbb{R}^d)$.

Define

$$\tilde{\mathcal{D}} := \operatorname{span}\{f_1 \otimes f_2 \ :\ f_1, f_2 \in C_c^\infty(\mathbb{R}^d)\}. \tag{3.5.61}$$

Putting Lemma 3.5.48 and Lemma 3.5.47 together, we see that the mappings $\{\nu_t^{12},\ t \in [0,T]\}$, $\{\nu_t^{11},\ t \in [0,T]\}$, and $\{\nu_t^{22},\ t \in [0,T]\}$, defined at (3.4.31) and (3.4.33), are $\mathcal{M}^+(\mathbb{R}^{2d})$-valued solutions of the evolution equation for $(\bar{A} + \bar{h}, \tilde{\mathcal{D}})$, that is,

$$\nu_t^{12} f = \nu_0^{12} f + \int_0^t \nu_u^{12}(\bar{A}f + \bar{h}f)\, du, \quad \forall\, t \in [0,T],\ \forall f \in \tilde{\mathcal{D}}, \tag{3.5.62}$$

with identical relations for $\nu_\cdot^{11}$ and $\nu_\cdot^{22}$ in place of $\nu_\cdot^{12}$. In order to prove Lemma 3.4.43, it remains to show that (3.5.62) holds not only for $f \in \tilde{\mathcal{D}}$, but for each f in the larger domain $\operatorname{span}\{1, C_c^\infty(\mathbb{R}^{2d})\}$. That is, it must be shown that the mappings $\{\nu_t^{12},\ t \in [0,T]\}$, $\{\nu_t^{11},\ t \in [0,T]\}$, and $\{\nu_t^{22},\ t \in [0,T]\}$ are $\mathcal{M}^+(\mathbb{R}^{2d})$-valued solutions of the evolution equation for $(\bar{A} + \bar{h}, \operatorname{span}\{1, C_c^\infty(\mathbb{R}^{2d})\})$. To this end we need the following result, whose proof is deferred to Appendix A.1:

Lemma 3.5.49. *Suppose Conditions 3.3.31–3.3.33 hold. Then the closure of the relation $\{(f, \bar{A}f) : f \in \tilde{\mathcal{D}}\}$ in the supremum norm of $B(\mathbb{R}^{2d}) \times B(\mathbb{R}^{2d})$ contains the relation $\{(f, \bar{A}f) : f \in C_c^\infty(\mathbb{R}^{2d})\}$.*

From Lemma 3.5.49 and the notions of bp-closedness and bp-closure of a relation (see foot of page 111 of Ethier and Kurtz [8]), we see that

$$\{(f, \bar{A}f) : f \in C_c^\infty(\mathbb{R}^{2d})\} \subset \text{bp-closure}\{(f, \bar{A}f) : f \in \tilde{\mathcal{D}}\}. \tag{3.5.63}$$

Now put

$$S^{12} := \left\{ (f, g) \in B(\mathbb{R}^{2d}) \times B(\mathbb{R}^{2d}) : \nu_t^{12} f = \nu_0^{12} f + \int_0^t \nu_s^{12}(g + \bar{h}f)\, ds,\ \forall t \in [0, T] \right\}, \tag{3.5.64}$$

and observe that S^{12} is a linear subspace of $B(\mathbb{R}^{2d}) \times B(\mathbb{R}^{2d})$. By (3.4.32) we have

$$\sup_{0 \le t \le T} \nu_t^{12}(\mathbb{R}^{2d}) < \infty,$$

and therefore, since $\bar{h} \in B(\mathbb{R}^{2d})$, it follows from the dominated convergence theorem that S^{12} is bp-closed in $B(\mathbb{R}^{2d}) \times B(\mathbb{R}^{2d})$. Since the $\mathcal{M}^+(\mathbb{R}^{2d})$-valued mapping $\{\nu_t^{12},\ t \in [0, T]\}$ solves the evolution equation for $(\bar{A} + \bar{h}, \tilde{\mathcal{D}})$, we have $\{(f, \bar{A}f) : f \in \tilde{\mathcal{D}}\} \subset S^{12}$, and therefore, from the bp-closedness of S^{12} and (3.5.63), we have

$$\{(f, \bar{A}f) : f \in C_c^\infty(\mathbb{R}^{2d})\} \subset S^{12}. \tag{3.5.65}$$

Next, observe from (3.4.42) and Lemma D.3.146 that the operator $(\bar{A}, C_c^\infty(\mathbb{R}^{2d}))$ is conservative (see Definition D.2.139), that is,

$$(1, 0) \in \text{bp-closure}\{(f, \bar{A}f) : f \in C_c^\infty(\mathbb{R}^{2d})\}. \tag{3.5.66}$$

In the light of (3.5.66), (3.5.65), and the bp-closedness of S^{12}, we then get

$$(1, 0) \in S^{12}. \tag{3.5.67}$$

Now (3.5.65) and (3.5.67), together with the linearity of S^{12} show that

$$\{(f, \bar{A}f) : f \in \text{span}\{1, C_c^\infty(\mathbb{R}^{2d})\}\} \subset S^{12},$$

which, in view of (3.5.64), shows that $\{\nu_t^{12},\ t \in [0, T]\}$ is an $\mathcal{M}^+(\mathbb{R}^{2d})$-valued solution of the evolution equation for $(\bar{A} + \bar{h}, \text{span}\{1, C_c^\infty(\mathbb{R}^{2d})\})$. Defining S^{11} as in

(3.5.64), but with ν^{11} in place of ν^{12}, we can similarly show that $\{\nu_t^{11},\ t \in [0,T]\}$ is an $\mathcal{M}^+(\mathbb{R}^{2d})$-valued solution of the evolution equation for $(\bar{A}+\bar{h}, \text{span}\{1, C_c^\infty(\mathbb{R}^{2d})\})$, and likewise for $\{\nu_t^{22},\ t \in [0,T]\}$. $\square$

Chapter 4

A Result on Singularly Perturbed Itô SDEs

In Section 1.1 we briefly summarized a result of Blankenship and Papanicolaou [5] which says that, subject to certain conditions, the law of the solution $\{X_t^\epsilon\}$ of a singularly perturbed ODE of the form (1.1.5) converges weakly to the law of the solution $\{\bar{X}_t\}$ of an SDE of the form (1.1.4) as $\epsilon \to 0$. Our goal in the present chapter is to extend this result to singularly perturbed SDE's of the form (1.1.14). We formulate conditions on (1.1.14) which are sufficient to ensure that the law of the solution $\{X_t^\epsilon\}$ converges weakly to the law of the solution $\{\bar{X}_t\}$ of an SDE of the form (1.1.15), provided that the coefficients $\bar{b}(\cdot)$ and $\bar{c}(\cdot)$ in (1.1.15) are given by formulae that will be established in this chapter. This result is a necessary preliminary in attaining the second major goal of the thesis, namely establishing convergence of nonlinear filters, which is undertaken in the next chapter. In addition, in the present chapter we highlight and make explicit the construction of so-called "perturbed test functions". These are a necessary tool in establishing the convergence result of this chapter, and will also be essential for dealing with convergence of nonlinear filters in Chapter 5.

4.1 Problem Formulation and Conditions

In the present chapter we study a *singularly perturbed SDE* of the form

$$dX_t^\epsilon = \frac{1}{\epsilon}F(X_t^\epsilon, Z_t^\epsilon)\,dt + G(X_t^\epsilon, Z_t^\epsilon)\,dt + B(X_t^\epsilon, Z_t^\epsilon)\,dW_t, \quad X_0^\epsilon = x_0 \in \mathbb{R}^d, \qquad (4.1.1)$$

parametrized by $\epsilon \in (0,1]$, where $\{Z_t^\epsilon\}$ is a "fast" perturbing process defined by the "rescaling"

$$Z_t^\epsilon := Z_{t/\epsilon^2}, \ \forall t \in [0,\infty),$$

of a given process $\{Z_t\}$, which will be assumed to have certain regularity and ergodic properties that will soon be precisely formulated (in Conditions 4.1.52 and 4.1.54 to follow). The SDE (4.1.1) is essentially the relation (1.1.14) in terms of which we formulated our research goals in Chapter 1, but is slightly more general in that we here allow the function B to depend not just on the system state $\{X_t^\epsilon\}$, but also on the fast perturbing process $\{Z_t^\epsilon\}$ as well. We allow this additional generality not because it is really meaningful from the modeling point of view (which requires that B need only depend on X_t^ϵ) but simply because it is almost as easy to analyze this case as the more special one in which B depends on X_t^ϵ only. Our goal is to characterize the limit of $\{X_t^\epsilon\}$ as $\epsilon \to 0$. To this end we postulate the following conditions for the SDE (4.1.1):

Condition 4.1.50. *$\{W_t,\ t \in [0,\infty)\}$ is a standard $\mathbb{R}^r$-valued Wiener process on the complete probability space $(\Omega, \mathcal{F}, P)$.*

Condition 4.1.51. *S is a compact metric space, and the mappings $F : \mathbb{R}^d \times S \to \mathbb{R}^d$, $G : \mathbb{R}^d \times S \to \mathbb{R}^d$, and $B : \mathbb{R}^d \times S \to \mathbb{R}^{d\times r}$ satisfy*

$$F^i \in C^{3,0}(\mathbb{R}^d \times S),\ G^i \in C^{2,0}(\mathbb{R}^d \times S),\ B^{i,j} \in C^{2,0}(\mathbb{R}^d \times S), \quad 1 \le i \le d,\ 1 \le j \le r,$$

and have uniformly bounded first x-derivatives, namely

$$\sup_{\mathbb{R}^d \times S} \left(|\partial_k F^i(x,z)| + |\partial_k G^i(x,z)| + |\partial_k B^{i,j}(x,z)| \right) < \infty, \ 1 \le i,k \le d, \ 1 \le j \le r. \tag{4.1.2}$$

Condition 4.1.52. *There is an S-valued Markov process $\{Z_t,\ t \in [0,\infty)\}$ defined on the probability space $(\Omega, \mathcal{F}, P)$ with a transition probability function $P_t(z,\Gamma)$ and initial distribution $\mu_0 \in \mathcal{P}(S)$, and the processes $\{Z_t\}$ and $\{W_t\}$ are independent. For each $\epsilon \in (0,1]$ the process $\{Z_t^\epsilon,\ t \in [0,\infty)\}$ in (4.1.1) is given by*

$$Z_t^\epsilon := Z_{t/\epsilon^2},\ \forall t \in [0,\infty). \tag{4.1.3}$$

Furthermore, the transition probability $P_t(z,\Gamma)$ of the Markov process $\{Z_t\}$ is such that, for each $\Psi \in C(S)$ and $t \in [0,\infty)$, the mapping

$$T_t\Psi(\cdot) := \int_S \Psi(z')\, P_t(\cdot, dz'), \tag{4.1.4}$$

is a member of $C(S)$ and defines a conservative (i.e. $T_t 1 \equiv 1$) positive strongly continuous contraction semigroup on $C(S)$ with infinitesimal generator denoted by $(\mathfrak{Q}, \mathcal{D}(\mathfrak{Q}))$.

Remark 4.1.53. In the terminology of Ethier and Kurtz ([8], page 166) the semigroup $\{T_t\}$ is a *Feller semigroup*. By the compactness of S, together with standard results on Feller semigroups (see, for example, Proposition 4.2.4 of Ethier and Kurtz [8]), without loss of generality we shall assume that the sample paths of $\{Z_t\}$ are corlol.

Condition 4.1.54 (Geometric ergodicity of $\{Z_t\}$). *There exists some unique $\bar{P} \in \mathcal{P}(S)$ which is invariant for the transition probability $P_t(z,\Gamma)$, namely*

$$\bar{P}(\Gamma) = \int_S P_t(z,\Gamma)\, \bar{P}(dz), \quad \forall\, t \in [0,\infty), \quad \forall\, \Gamma \in \mathcal{B}(S). \tag{4.1.5}$$

Moreover, for this invariant distribution $\bar{P}$ we have

$$\int_0^\infty \|T_t\Psi - \bar{P}\Psi\|\, dt < \infty, \quad \forall \Psi \in C(S). \tag{4.1.6}$$

Condition 4.1.55. *The probability measure $\bar{P}$ in Condition 4.1.54 is such that*

$$\bar{P}F^i(x,\cdot) := \int_S F^i(x,z)\,\bar{P}(dz) = 0, \quad \forall x \in \mathbb{R}^d,\ 1 \le i \le d. \tag{4.1.7}$$

The following result, which is suggested by Remark 12.2.3 of Ethier and Kurtz [8], lists some useful consequences of Condition 4.1.54, which will be needed to establish the asymptotic properties of $\{X_t^\epsilon\}$. The proof is given in Appendix B.

Lemma 4.1.56. *Suppose that Condition 4.1.54 holds. Then*

(i) for each $z \in S$ there exists a signed Borel measure $\chi(z,\cdot)$ on S such that $\sup_z \|\chi(z,\cdot)\|_{TV} < \infty$, and such that for every $\Psi \in C(S)$ we have

$$\int_0^\infty [T_t\Psi(\cdot) - \bar{P}\Psi]\,dt = \int_S \Psi(z')\chi(\cdot,dz') \in C(S); \tag{4.1.8}$$

(ii) for $\Psi \in C(S)$ such that

$$\int_S \Psi(z)\,\bar{P}(dz) = 0,$$

the function

$$\Phi(z) := \int_S \Psi(z')\chi(z,dz'), \quad z \in S,$$

belongs to $\mathcal{D}(\Omega)$ and solves the "Poisson equation "

$$\Omega\Phi = -\Psi,$$

where $(\Omega, \mathcal{D}(\Omega))$ is the infinitesimal generator of the Markov process $\{Z_t\}$ (recall Condition 4.1.52).

(iii) for every $g \in C^{1,0}(\mathbb{R}^d \times S)$, the function

$$f(x,z) := \int_S g(x,z')\,\chi(z,dz'), \quad (x,z) \in (\mathbb{R}^d \times S)$$

belongs to $C^{1,0}(\mathbb{R}^d \times S)$, and

$$\partial_j f(x,z) = \int_S \partial_j g(x,z')\,\chi(z,dz'), \quad (x,z) \in \mathbb{R}^d \times S,\ 1 \le j \le d.$$

Remark 4.1.57. The left-hand integral in (4.1.8) is a Riemann integral in the Banach space $C(S)$ of the mapping

$$t \in [0,\infty) \longmapsto [T_t\Psi - \bar{P}\Psi] \in C(S),$$

which is continuous since $\{T_t\}$ is strongly continuous. Existence of this integral as a member of $C(S)$ follows from Condition 4.1.54 and Lemma 1.1.4 of Ethier and Kurtz [8].

Remark 4.1.58. Condition 4.1.51, together with standard existence and uniqueness results for stochastic differential equations (see, for example, Theorem 5.1.1 of Kallianpur [15]), ensures that (4.1.1) has a pathwise unique strong solution $\{X_t^\epsilon,\ t \in [0,\infty)\}$ adapted to the filtration $\{\mathcal{F}_t^{W,Z^\epsilon}\}$ defined by

$$\mathcal{F}_t^{W,Z^\epsilon} := \sigma\{W_s,\ Z_s^\epsilon,\ s \in [0,t]\} \vee \mathcal{N}(P), \tag{4.1.9}$$

for each $\epsilon \in (0,1]$.

Our goal is to show that $\{X_t^\epsilon\}$ converges weakly to a diffusion process as $\epsilon \to 0+$, and to characterize this diffusion. To this end, we use the signed Borel measure $\chi(z,\cdot)$ from Lemma 4.1.56(i) to define the coefficients

$$b^j(x) := \int_S \left\{ G^j(x,z) + \int_S \sum_{i=1}^d \partial_i F^j(x,z')\chi(z,dz')F^i(x,z) \right\} \bar{P}(dz), \quad x \in \mathbb{R}^d, \tag{4.1.10}$$

$$\begin{aligned} a^{ij}(x) := \int_S \Big\{ \int_S F^i(x,z')\chi(z,dz')F^j(x,z) + \int_S F^j(x,z')\chi(z,dz')F^i(x,z) \\ + [BB^T(x,z)]^{ij} \Big\} \bar{P}(dz), \quad x \in \mathbb{R}^d. \end{aligned} \tag{4.1.11}$$

Also define the diffusion operator $(\mathcal{L}, \mathcal{D}(\mathcal{L}))$ by

$$\begin{aligned} &\mathcal{D}(\mathcal{L}) := C_c^\infty(\mathbb{R}^d), \\ &\mathcal{L}\phi(x) := \sum_{i=1}^d b^i(x)\partial_i\phi(x) + \frac{1}{2}\sum_{i,j=1}^d a^{ij}(x)\partial_i\partial_j\phi(x), \quad x \in \mathbb{R}^d,\ \phi \in \mathcal{D}(\mathcal{L}). \end{aligned} \tag{4.1.12}$$

Remark 4.1.59. The main result of this section (Theorem 4.2.63) is that $\{X_t^\epsilon\}$ converges weakly to a diffusion process with second order differential operator $\mathcal{L}$. For this, it is necessary that the matrix $a(x) \in \mathbb{R}^{d\times d}$ in (4.1.11) be nonnegative definite for each $x \in \mathbb{R}^d$. To check that this is indeed the case, note that it suffices to show the same for the $d \times d$ matrix

$$\hat{a}^{ij}(x) := \int_S \int_S \left[F^i(x,z')\chi(z,dz')F^j(x,z) + F^j(x,z')\chi(z,dz')F^i(x,z)\right] \bar{P}(dz). \tag{4.1.13}$$

To this end, since $F^i(x,\cdot) \in C(S)$ we observe from (4.1.4) that

$$T_t F^i(x,\cdot)(z) = \int_S F^i(x,z')P_t(z,dz'), \quad \forall z \in S,$$

defines a member of $C(S)$ for each $x \in \mathbb{R}^d$. Also, from (4.1.8) and Condition 4.1.55, we have

$$\int_0^\infty T_t F^i(x,\cdot)(z)\,dt = \int_S F^i(x,z')\,\chi(z,dz'), \quad \forall z \in S.$$

Thus

$$\hat{a}^{ij}(x) = \int_S \left[\int_0^\infty T_t F^i(x,\cdot)(z) F^j(x,z)\,dt + \int_0^\infty T_t F^j(x,\cdot)(z) F^i(x,z)\,dt\right] \bar{P}(dz),$$

which, by $\bar{P}F^i(x,\cdot) = 0$ (recall (4.1.7)) and Fubini's theorem (recall (4.1.6)), can be written as

$$\hat{a}^{ij}(x) = \int_0^\infty \mathrm{E}[F^i(x,\hat{Z}_t)F^j(x,\hat{Z}_0) + F^j(x,\hat{Z}_t)F^i(x,\hat{Z}_0)]\,dt, \tag{4.1.14}$$

where $\{\tilde{Z}_t\}$ is a stationary process corresponding to the transition probability $P_t(z,\Gamma)$ and initial measure $\bar{P}$. This can further be expressed as

$$\hat{a}^{ij}(x) = \int_{-\infty}^\infty \mathrm{E}[F^i(x,\hat{Z}_t)F^j(x,\hat{Z}_0)]\,dt, \quad x \in \mathbb{R}^d.$$

The above expression is the spectral density matrix of a stationary process evaluated at zero, and is therefore nonnegative definite.

Remark 4.1.60. From (4.1.2), (4.1.10), (4.1.11), and Lemma 4.1.56(i) it follows that $b^i(\cdot)$ and $a^{ij}(\cdot)$ are continuous functions on $\mathbb{R}^d$, and that there exists a constant $C \in [0, \infty)$ such that

$$|b^i(x)| \leq C[1 + |x|], \quad |a^{ij}(x)| \leq C[1 + |x|^2], \quad \forall x \in \mathbb{R}^d, \ 1 \leq i, j \leq d. \tag{4.1.15}$$

Therefore, by Lemma D.3.146, we know that $\mathfrak{L}$ is a *conservative* linear operator (see Definition D.2.139). By Theorem 5.3.10 and Proposition 5.3.5 of Ethier and Kurtz [8] it also follows that for every $x \in \mathbb{R}^d$ there exist continuous solutions of the martingale problem for $(\mathfrak{L}, \delta_x)$. The next condition strengthens this and insists that there is exactly *one* such solution:

Condition 4.1.61. *The martingale problem for $(\mathfrak{L}, \delta_x)$ is well posed for each $x \in \mathbb{R}^d$ (recall that $\mathfrak{L}$ is defined by* (4.1.10), (4.1.11), *and* (4.1.12)*).*

Remark 4.1.62. The preceding Conditions 4.1.50, 4.1.51, 4.1.52, 4.1.54, 4.1.55, and 4.1.61 will usually be invoked together, and we will therefore refer to these conditions collectively as Condition **AI**. When specialized to the case $B \equiv 0$ in (4.1.1), these are essentially the conditions adopted by Blankenship and Papanicolaou [5].

4.2 Main Result

The following result is a generalization of the Theorem on page 449 of Blankenship and Papanicolaou [5], to which it reduces for $B \equiv 0$ in (4.1.1):

Theorem 4.2.63. *Suppose Condition* **AI**, *and let P_0 be the (unique) element in $\mathcal{P}(C_{\mathbb{R}^d}[0, \infty))$ solving the martingale problem for $(\mathfrak{L}, \delta_{x_0})$. Then*

$$\lim_{\epsilon \to 0+} \mathcal{L}(X^\epsilon) = P_0 \quad \textit{in} \quad \mathcal{P}(C_{\mathbb{R}^d}[0, \infty)). \tag{4.2.16}$$

Remark 4.2.64. The expression for the operator $\mathcal{L}$ in (4.1.12) is very similar to that of the generator of the limiting diffusion in Blankenship and Papanicolaou (see (4.11) on page 448 of [5]), the only difference being the presence of the last term under the integral sign in (4.1.11), which corresponds to the "averaging" of the B matrix specifying the stochastic integral in (4.1.1).

Remark 4.2.65. In proving Theorem 4.2.63 we follow the well-known method of first establishing relative compactness of a sequence of probability measure $\{\mathcal{L}(X^{\epsilon_n})\}$ for an arbitrary sequence $(\epsilon_n)_{n=1}^{\infty} \subset (0,1]$ converging to zero, and then showing that every weak limit of this sequence solves the well-posed martingale problem for $(\mathcal{L}, \delta_{x_0})$. In implementing this scheme, we shall use the method of Blankenship and Papanicolaou [5], which itself is motivated by an abstract approach to averaging problems due to Kurtz [19]. In rough outline the idea of the method is as follows: Let $\mathcal{C}^\epsilon$ be the operator for the martingale problem which is satisfied by the process $\{(X_t^\epsilon, Z_t^\epsilon)\}$ - this will be made precise in (4.2.20) and Proposition 4.2.67. For each $\phi \in \mathcal{D}(\mathcal{L})$ one must cleverly construct a uniformly bounded mapping $\psi^\phi : \mathbb{R}^d \times S \times (0,1] \to \mathbb{R}$ such that $f^{\epsilon,\phi}$ defined by

$$f^{\epsilon,\phi}(x,z) = \phi(x) + \epsilon\psi^\phi(x,z,\epsilon), \tag{4.2.17}$$

is in the domain of $\mathcal{C}^\epsilon$, and such that $\mathcal{C}^\epsilon f^{\epsilon,\phi}$ has the form

$$\mathcal{C}^\epsilon f^{\epsilon,\phi}(x,z) = \mathcal{L}\phi(x) + \epsilon\gamma^\phi(x,z,\epsilon), \tag{4.2.18}$$

for some uniformly bounded function $\gamma^\phi : \mathbb{R}^d \times S \times (0,1] \to \mathbb{R}$. One can think of the second term on the right side of (4.2.17) as a small perturbation of the "test function" ϕ, giving the "perturbed test function" $f^{\epsilon,\phi}$. This perturbation is carefully constructed to ensure that the result of $\mathcal{C}^\epsilon$ operating on $f^{\epsilon,\phi}$ is an equally small perturbation of $\mathcal{L}\phi$ (see (4.2.18)). Thus, for each $\phi \in \mathcal{D}(\mathcal{L})$, it follows that $\mathcal{C}^\epsilon f^{\epsilon,\phi}$ converges to $\mathcal{L}\phi$, uniformly in both z and x, as $\epsilon \to 0$. As will be seen, this convergence is the key to establishing both the weak relative compactness of $\mathcal{L}(X^{\epsilon_n})$, as well as the fact that each weak limit of this sequence solves the martingale problem for $(\mathfrak{L}, \delta_{x_0})$. The actual construction of the perturbed test functions is accomplished by Proposition 4.2.68, and depends crucially on the geometric ergodicity postulated for $\{Z_t\}$ by Condition 4.1.54.

Remark 4.2.66. For obvious reasons the strategy outlined in Remark 4.2.65 is called the *method of perturbed test functions.* This general approach was first given a heuristic formulation in the early works of Stratonovich [34]. An abstract and mathematically rigorous formulation was later established by Kurtz [19], and then used by Blankenship and Papanicolaou [5] to study the asymptotic properties of (4.1.1) in the case where $B(\cdot) \equiv 0$. The book of Kushner [24] and the paper of Kurtz [20] give several other examples and applications of this approach.

As a first step in implementing this approach, we must give a precise formulation of the operator $\mathcal{C}^\epsilon$ in (4.2.18). Recalling the infinitesimal generator $(\mathcal{Q}, \mathcal{D}(\mathcal{Q}))$ of the Markov process $\{Z_t\}$ in Condition 4.1.52, let

$$\begin{aligned}\tilde{\mathcal{D}} := \big\{ f \in C_c^{2,0}(\mathbb{R}^d \times S) \ : \ & f(x,\cdot) \in \mathcal{D}(\mathcal{Q}) \ \ \forall x \in \mathbb{R}^d, \text{ and the mapping} \\ & (x,z) \in \mathbb{R}^d \times S \mapsto \mathcal{Q}[f(x,\cdot)](z) \in \mathbb{R} \quad \text{belongs to} \quad C_c^{2,0}(\mathbb{R}^d \times S)\big\}.\end{aligned} \tag{4.2.19}$$

For each $\epsilon \in (0,1]$ define operator $\mathcal{C}^\epsilon$ with domain $\tilde{\mathcal{D}}$ as follows:

$$\begin{aligned}\mathcal{C}^\epsilon f(x,z) := {} & \frac{1}{\epsilon^2}\mathcal{Q}[f(x,\cdot)](z) + \sum_{i=1}^d \left[\frac{1}{\epsilon}F^i(x,z) + G^i(x,z)\right]\partial_i f(x,z) \\ & + \frac{1}{2}\sum_{i,j=1}^d [BB^T(x,z)]^{ij}\partial_i\partial_j f(x,z), \quad \forall (x,z) \in \mathbb{R}^d \times S, \ \forall f \in \tilde{\mathcal{D}}.\end{aligned} \tag{4.2.20}$$

Note that $\tilde{\mathcal{D}}$ is a linear subspace of $C_c^{2,0}(\mathbb{R}^d \times S)$, and that $\mathcal{C}^\epsilon f$ is an element of $C_c(\mathbb{R}^d \times S)$ for each $f \in \tilde{\mathcal{D}}$, so that $(\mathcal{C}^\epsilon, \tilde{\mathcal{D}})$ is a linear operator on $C_c(\mathbb{R}^d \times S)$ for every $\epsilon \in (0,1]$. The next proposition, whose proof is given in Appendix B, establishes that $\{(X_t^\epsilon, Z_t^\epsilon)\}$ solves the martingale problem for the operator $(\mathcal{C}^\epsilon, \tilde{\mathcal{D}})$ (see Definition D.2.134):

Proposition 4.2.67. *Suppose Condition* **AI** *(see Remark 4.1.62). For every $\epsilon \in (0,1]$ and $f \in \tilde{\mathcal{D}}$ the process*

$$M_t^{\epsilon,f} := f(X_t^\epsilon, Z_t^\epsilon) - \int_0^t \mathcal{C}^\epsilon f(X_s^\epsilon, Z_s^\epsilon)\, ds, \quad t \in [0,\infty) \tag{4.2.21}$$

is an $\{\mathcal{F}_t^{W,Z^\epsilon}\}$-martingale.

The next result, which is also proved in Appendix B, establishes that there do indeed exist "perturbed test functions" $f^{\epsilon,\phi}$ of the general form (4.2.17) such that (4.2.18) holds:

Proposition 4.2.68. *Suppose Condition* **AI** *(see Remark 4.1.62). Then, for each $\phi \in C_c^\infty(\mathbb{R}^d)$, there exist functions $f_1^\phi \in C_c^{3,0}(\mathbb{R}^d \times S)$, $f_2^\phi \in C_c^{2,0}(\mathbb{R}^d \times S)$, $f_3^\phi \in C_c^{2,0}(\mathbb{R}^d \times S)$, and $\gamma^\phi \in \bar{C}(\mathbb{R}^d \times S \times (0,1])$, with the following property: If $f^{\epsilon,\phi}(\cdot)$ is defined for each $\epsilon \in (0,1]$ by*

$$f^{\epsilon,\phi}(x,z) := \phi(x) + \epsilon f_1^\phi(x,z) + \epsilon^2 f_2^\phi(x,z) + \epsilon^2 f_3^\phi(x,z), \quad \forall\, (x,z) \in \mathbb{R}^d \times S, \tag{4.2.22}$$

then $f^{\epsilon,\phi} \in \tilde{\mathcal{D}}$, with

$$\mathcal{C}^\epsilon f^{\epsilon,\phi}(x,z) = \mathcal{L}\phi(x) + \epsilon\gamma^\phi(x,z,\epsilon), \quad \forall\, (x,z,\epsilon) \in \mathbb{R}^d \times S \times (0,1]. \tag{4.2.23}$$

Remark 4.2.69. The functions f_i^ϕ, $i = 1,2,3$, and $f^{\epsilon,\phi}$ provided by Proposition 4.2.68 will be used in the next section to establish Theorem 4.2.63. These functions will also play an essential role in the next chapter on convergence of nonlinear filters.

4.3 Proof of Theorem 4.2.63

The proof is given in two steps:

Step 1: Fix some sequence $(\epsilon_n)_{n=1}^{\infty} \subset (0,1]$ such that $\lim_{n\to\infty} \epsilon_n = 0$. We first establish that the sequence $\{\mathcal{L}(X^{\epsilon_n}),\ n \in \mathbb{N}\}$ is relatively compact in $\mathcal{P}(C_{\mathbb{R}^{d*}}[0,\infty))$ (note that we regard the $\mathbb{R}^d$-valued continuous process $\{X_t^{\epsilon}\}$ as an $\mathbb{R}^{d*}$-valued continuous process, where $\mathbb{R}^{d*}$ denotes the usual one-point (Alexandrov) compactification of $\mathbb{R}^d$, and therefore $\mathcal{L}(X^{\epsilon_n})$ is a probability measure in the Polish space $C_{\mathbb{R}^{d*}}[0,\infty)$). To this end, define the linear operator $(\mathcal{L}^*, \mathcal{D}(\mathcal{L}^*))$ on $C(\mathbb{R}^{d*})$ as follows:

$$\mathcal{D}(\mathcal{L}^*) := \{\phi \in C(\mathbb{R}^{d*}) : \phi|_{\mathbb{R}^d} - \phi(\Delta) \in C_c^{\infty}(\mathbb{R}^d)\},$$

$$\mathcal{L}^*\phi(x) := \begin{cases} \mathcal{L}(\phi|_{\mathbb{R}^d} - \phi(\Delta))(x), & \text{if } x \in \mathbb{R}^d; \\ 0, & \text{if } x = \Delta, \end{cases} \tag{4.3.24}$$

where the operator $(\mathcal{L}, \mathcal{D}(\mathcal{L}))$ is defined in (4.1.12), Δ denotes the point at infinity in the one-point compactification $\mathbb{R}^{d*}$, and $\phi|_{\mathbb{R}^d}$ indicates the restriction of ϕ to the domain $\mathbb{R}^d$. To simplify the notation, we shall write "n" instead of "ϵ_n" in superscripts, so that X^n is short for X^{ϵ_n}, $f^{n,\phi}$ is short for $f^{\epsilon_n,\phi}$, and so on. We are now ready to prove relative compactness of $\{\mathcal{L}(X^n),\ n \in \mathbb{N}\}$ in $\mathcal{P}(C_{\mathbb{R}^{d*}}[0,\infty))$. Fix some $\phi \in \mathcal{D}(\mathcal{L}^*)$, some $T \in [0,\infty)$, and define $\hat{\phi} \in C_c^{\infty}(\mathbb{R}^d)$ by

$$\hat{\phi} \equiv \phi|_{\mathbb{R}^d} - \phi(\Delta). \tag{4.3.25}$$

Also, for each $n \in \mathbb{N}$, define

$$U_n(t) := f^{n,\hat{\phi}}(X_t^n, Z_t^n) + \phi(\Delta), \quad t \in [0,\infty); \tag{4.3.26a}$$

$$V_n(t) := \mathfrak{C}^n f^{n,\hat{\phi}}(X_t^n, Z_t^n), \quad t \in [0,\infty), \tag{4.3.26b}$$

where $f^{n,\hat{\phi}} := f^{\epsilon_n,\hat{\phi}} \in \tilde{\mathcal{D}}$ is given by (4.2.22) in terms of the functions $f_i^{\hat{\phi}}$, $i = 1,2,3$, provided by Proposition 4.2.68. By Proposition 4.2.67,

$$U_n(t) - \int_0^t V_n(s)\,ds \quad \text{is an } \{\mathcal{F}_t^{W,Z^n}\}\text{-martingale} \quad \forall n \in \mathbb{N}, \tag{4.3.27}$$

and, by (4.2.22) and (4.3.26a), we have

$$U_n(t) = \phi(\Delta) + \hat{\phi}(X_t^n) + \epsilon_n f_1^{\hat{\phi}}(X_t^n, Z_t^n) + \epsilon_n^2[f_2^{\hat{\phi}}(X_t^n, Z_t^n) + f_3^{\hat{\phi},n}(X_t^n, Z_t^n)], \quad t \in [0,\infty). \tag{4.3.28}$$

From Proposition 4.2.68 we see that the $f_i^{\hat{\phi}}$ are uniformly bounded over $\mathbb{R}^d \times S$ for each $i = 1,2,3$. Thus, from (4.3.28), (4.3.25), and the fact that X_t^n has values in $\mathbb{R}^d$ (and never takes the value Δ), we see that

$$\operatorname*{bp-lim}_{n\to\infty} \sup_{t\in[0,T]} |U_n(t) - \phi(X_t^n)| = 0. \tag{4.3.29}$$

Also, by (4.2.23), (4.3.24), and (4.3.26b) we have

$$\begin{aligned} V_n(t) =& \mathcal{L}\hat{\phi}(X_t^n) + \epsilon_n \gamma^{\hat{\phi}}(X_t^n, Z_t^n, \epsilon_n) \\ =& \mathcal{L}^*\phi(X_t^n) + \epsilon_n \gamma^{\hat{\phi}}(X_t^n, Z_t^n, \epsilon_n), \quad t \in [0,\infty), \end{aligned} \tag{4.3.30}$$

whence, by the uniform boundedness of $\gamma^{\hat{\phi}}$, (see Proposition 4.2.68)

$$\operatorname*{bp-lim}_{n\to\infty} \sup_{t\in[0,T]} |V_n(t) - \mathcal{L}^*\phi(X_t^n)| = 0. \tag{4.3.31}$$

Since $\mathcal{L}^*\phi \in \bar{C}(\mathbb{R}^{d*})$, and thus $\mathcal{L}^*\phi$ is uniformly bounded over $\mathbb{R}^d$, we see from (4.3.29) and (4.3.31) that

$$\lim_{n\to\infty} \mathrm{E}\left[\sup_{t\in[0,T]} |U_n(t) - \phi(X_t^n)|\right] = 0, \tag{4.3.32}$$

$$\limsup_{n\to\infty} \mathrm{E}\left[\sup_{t\in[0,T]} |V_n(t)|\right] < \infty. \tag{4.3.33}$$

Now $\mathcal{D}(\mathcal{L}^*)$ is an algebra in $\bar{C}(\mathbb{R}^{d*})$. Hence, taking $E := \mathbb{R}^{d*}$, $\mathcal{F}_t^n := \mathcal{F}_t^{W,Z^n}$, and $C_a := \mathcal{D}(\mathcal{L}^*)$ in Theorem D.1.133, and using (4.3.27), (4.3.32), (4.3.33), we see that

$\{\mathcal{L}(\phi(X^n)),\ n \in \mathbb{N}\}$ is relatively compact in $\mathcal{P}(C_{\mathbb{R}}[0,\infty))$ for every $\phi \in \mathcal{D}(\mathcal{L}^*)$. But the Stone-Weierstrass theorem ensures that $\mathcal{D}(\mathcal{L}^*)$ is dense in $\bar{C}(\mathbb{R}^{d*})$, and hence we can take $E := \mathbb{R}^{d*}$ and $U := \mathcal{D}(\mathcal{L}^*)$ in Theorem D.1.132 to conclude that $\{\mathcal{L}(X^n),\ n \in \mathbb{N}\}$ is relatively compact in $\mathcal{P}(C_{\mathbb{R}^{d*}}[0,\infty))$ (the compact containment condition required by Theorem D.1.132 holds trivially by the compactness of $E := \mathbb{R}^{d*}$).

Step 2: In this step we complete the proof. First we need the following:

Fact 4.3.70. *Let $\tilde{P} \in \mathcal{P}(C_{\mathbb{R}^{d*}}[0,\infty))$ be a solution of the martingale problem for $(\mathcal{L}^*, \delta_x)$, for some $x \in \mathbb{R}^d$. Then $\tilde{P}(C_{\mathbb{R}^d}[0,\infty)) = 1$, and, if P denotes the restriction of $\tilde{P}$ to $C_{\mathbb{R}^d}[0,\infty)$, then P is a solution of the martingale problem for $(\mathcal{L}, \delta_x)$.*

Proof of Fact 4.3.70. This is an immediate consequence of Theorem D.2.140, together with the fact that the operator $(\mathcal{L}, \mathcal{D}(\mathcal{L}))$ is conservative (see Remark 4.1.60).

□

Fix some $T \in [0,\infty)$ and some $t \in [0,T]$. Then, for finite subsets $(t_i)_{i=1}^k \subset [0,t]$ and $(h_i)_{i=1}^k \subset \bar{C}(\mathbb{R}^{d*})$, from (4.3.28) and (4.3.30) we have

$$\sup_n \sup_{t\in[0,T]} \mathrm{E}[|U_n(t)|] < \infty, \tag{4.3.34}$$

$$\sup_n \sup_{t\in[0,T]} \mathrm{E}[|V_n(t)|] < \infty, \tag{4.3.35}$$

$$\lim_{n\to\infty} \mathrm{E}\left[|U_n(t) - \phi(X_t^n)| \prod_{i=1}^k h_i(X_{t_i}^n)\right] = 0, \quad \forall \phi \in \mathcal{D}(\mathcal{L}^*), \tag{4.3.36}$$

$$\lim_{n\to\infty} \mathrm{E}\left[|V_n(t) - \mathcal{L}^*\phi(X_t^n)| \prod_{i=1}^k h_i(X_{t_i}^n)\right] = 0, \quad \forall \phi \in \mathcal{D}(\mathcal{L}^*). \tag{4.3.37}$$

In view of Fact 4.3.70 and Condition 4.1.61, it follows that uniqueness holds for the $C_{\mathbb{R}^{d*}}[0,\infty)$ martingale problem for $(\mathcal{L}^*, \delta_{x_0})$ (since $x_0 \in \mathbb{R}^d$). Thus, by (4.3.27) and (4.3.34) to (4.3.37), together with the relative compactness of $\{\mathcal{L}(X^n)\}$ in $\mathcal{P}(C_{\mathbb{R}^{d_*}}[0,\infty))$ (established in Step 1), we can use Theorem D.2.141 to see that there exists some solution $\tilde{P}_0 \in \mathcal{P}(C_{\mathbb{R}^{d_*}}[0,\infty))$ of the martingale problem for $(\mathcal{L}^*, \delta_{x_0})$ (see Remark D.2.136) such that

$$\lim_{n\to\infty} \mathcal{L}(X^n) = \tilde{P}_0 \quad \text{in} \quad \mathcal{P}(C_{\mathbb{R}^{d_*}}[0,\infty)). \tag{4.3.38}$$

By Fact 4.3.70 and Condition 4.1.61, we see that the restriction of $\tilde{P}_0$ to $C_{\mathbb{R}^d}[0,\infty)$ is the *unique* solution P_0 of the martingale problem for $(\mathcal{L}, \delta_{x_0})$, and hence the result follows from (4.3.38) and the fact that $\mathcal{L}(X^n)(C_{\mathbb{R}^d}[0,\infty)) = 1$, for every $n \in \mathbb{N}$. □

4.4 A Special Case

We conclude this chapter by considering the special case specified by the following condition.

Condition 4.4.71. *The mapping $B(\cdot)$ in* (4.1.1) *is a function of its first argument only, namely $B : \mathbb{R}^d \to \mathbb{R}^{d\times r}$.*

If Condition 4.4.71 holds, then, from (4.1.11) and (4.1.13), we can write

$$a(x) = \hat{a}(x) + BB^T(x) \quad x \in \mathbb{R}^d.$$

Let $c : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ be some Borel-measurable square root of the positive semidefinite matrix $\hat{a}(\cdot)$. Then the last relation can be written as

$$a(x) = \begin{bmatrix} c(x) & B(x) \end{bmatrix} \begin{bmatrix} c^T(x) \\ B^T(x) \end{bmatrix}, \quad x \in \mathbb{R}^d. \tag{4.4.39}$$

We now have the following result which will be needed for the next chapter.

Theorem 4.4.72. *Suppose Conditions **AI** and 4.4.71 hold (recall Remark 4.1.62), let $P_0 \in \mathcal{P}(C_{\mathbb{R}^d}[0,\infty))$ be a solution of the martingale problem for $(\mathcal{L}, \delta_{x_0})$, and let $\bar{X}_\cdot$ be a generic element of $C_{\mathbb{R}^d}[0,\infty)$. Then* *(i) there exists a complete probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$, which is an extension of the probability space $(C_{\mathbb{R}^d}[0,\infty), \mathcal{B}(C_{\mathbb{R}^d}[0,\infty)), P_0)$, and equipped with a standard filtration $\{\bar{\mathcal{F}}_t\}$ together with an $\mathbb{R}^{r+d}$-valued $\{\bar{\mathcal{F}}_t\}$-Wiener process $\{\hat{W}_t\} \equiv \{(\bar{W}_t, \bar{V}_t),\ t \in [0,\infty)\}$, such that $\{(\bar{\Omega}, \bar{\mathcal{F}}, \{\bar{\mathcal{F}}_t\}, \bar{P}), (\bar{X}_t, \hat{W}_t)\}$ is a weak solution of the SDE*

$$d\bar{X}_t^i = b^i(\bar{X}_t)\,dt + \sum_{j=1}^{r} B^{ij}(\bar{X}_t)\,d\bar{W}_t^j + \sum_{j=1}^{d} c^{ij}(\bar{X}_t)\,d\bar{V}_t^j, \quad 1 \le i \le d, \quad t \in [0,\infty), \tag{4.4.40}$$

with

$$\bar{X}_0 = x_0 \qquad a.s.; \tag{4.4.41}$$

(ii) $\lim_{\epsilon\to 0} \mathcal{L}(X^\epsilon) = \mathcal{L}(\bar{X})$ *in* $\mathcal{P}(C_{\mathbb{R}^d}[0,\infty))$.

Proof. (i) From (4.4.40) and Itô's formula one easily sees that the processes

$$M_t^i := \bar{X}_t^i - \bar{X}_t^i - \int_0^t b^i(\bar{X}_s)\,ds, \quad t \in [0,\infty)$$

and

$$M_t^{ik} := \bar{X}_t^i \bar{X}_t^k - \bar{X}_0^i \bar{X}_0^k - \int_0^t [\bar{X}_s^i\, b^k(\bar{X}_s) + \bar{X}_s^k\, b^i(\bar{X}_s) + a^{ik}(\bar{X}_s)]\,ds, \quad t \in [0,\infty)$$

are continuous $\{\mathcal{F}_{t+}^{\bar{X}}\}$-local martingales on $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$. Therefore, the result follows directly from Proposition 5.4.6 of Karatzas and Shreve [16].

(ii) This follows from (i) and Theorem 4.2.63. □

Note that, compared with the martingale problem representation of $\{\bar{X}_t\}$, the SDE (4.4.40) has one more element, namely the Wiener process $\{\hat{W}_t\}$. Its first component, $\{\bar{W}_t\}$, will be used in the formulation of the filtering problem of the next chapter.

Chapter 5

Convergence of Nonlinear Filters

In this chapter we address the second problem of the thesis and show weak convergence of nonlinear filters as outlined in Section 1.1 (these goals are also recalled in greater detail in Section 5.1 which follows). In brief, we shall introduce a martingale problem for the probability measure-valued solutions of the normalized filter equation, and use the uniqueness result of Chapter 3 to show that the martingale problem for the limiting nonlinear filter is well-posed. This will enable us to apply a powerful convergence theorem of Bhatt and Karandikar [3] to establish convergence of the nonlinear filters. An essential role will be played by the perturbed test functions furnished by Proposition 4.2.68, which will be used to verify some of the conditions of Bhatt and Karandikar's result.

Convergence of nonlinear filters has received some attention in the established literature, but we adopt a different approach to this problem from that used in previous works, and establish convergence subject to significantly weaker restrictions. We compare our results with these other works in Remark 5.3.101, Remark 5.3.103, and Remark 5.3.104, at the end of the chapter.

5.1 Introduction, Problem Formulation, and Main Result

We consider the SDE

$$dX_t^\epsilon = \frac{1}{\epsilon}F(X_t^\epsilon, Z_t^\epsilon)\,dt + G(X_t^\epsilon, Z_t^\epsilon)\,dt + B(X_t^\epsilon)\,dW_t, \quad X_0^\epsilon = x_0,\ t \in [0,T] \quad (5.1.1)$$

on the probability space $(\Omega, \mathcal{F}, P)$. This is essentially the SDE (4.1.1) whose asymptotics were studied in Chapter 4, except that B is assumed to be a function of X_t^ϵ only, rather than a function of $(X_t^\epsilon, Z_t^\epsilon)$, and we limit attention to a fixed time horizon $[0,T]$, for some $T \in (0,\infty)$. Suppose that the conditions imposed in Theorem 4.4.72 hold, so that, in particular, equation (5.1.1) has a unique solution. For each $\epsilon \in (0,1]$ we will take the *signal* to be the $\mathbb{R}^d$-valued process $\{X_t^\epsilon,\ t \in [0,T]\}$ given by (5.1.1), and define the corresponding $\mathbb{R}^r$-dimensional *observation process* $\{Y_t^\epsilon,\ t \in [0,T]\}$ by

$$Y_t^\epsilon := W_t + \int_0^t h(X_u^\epsilon)\,du, \quad t \in [0,T]. \quad (5.1.2)$$

Note that the Wiener process $\{W_t\}$ in the observation equation (5.1.2) also appears in the model dynamics (5.1.1), so that the signal process $\{X_t^\epsilon\}$ is effectively conditioned by the observation process $\{Y_t^\epsilon\}$. If we define the observation filtration by

$$\mathcal{F}_t^{Y^\epsilon} := \sigma\{Y_s^\epsilon,\ 0 \le s \le t\} \vee \mathcal{N}(P), \quad \epsilon \in (0,1], \quad (5.1.3)$$

then Lemma 2.3.5 yields a $\mathcal{P}(\mathbb{R}^d)$-valued process $\{\pi_t^\epsilon,\ t \in [0,T]\}$ which is $\{\mathcal{F}_{t+}^{Y^\epsilon}\}$-optional and satisfies

$$\pi_t^\epsilon f = \mathrm{E}[f(X_t^\epsilon) | \mathcal{F}_{t+}^{Y^\epsilon}] \quad a.s. \quad (5.1.4)$$

for each $t \in [0,T]$ and $f \in B(\mathbb{R}^d)$. This process is called the *nonlinear filter* for the signal process $\{X_t^\epsilon\}$ given the observation $\{Y_t^\epsilon\}$. In this chapter our goal is to study and characterize the asymptotic limit, in the sense of convergence in law, of the nonlinear filter $\{\pi_t^\epsilon,\ t \in [0,T]\}$ as $\epsilon \to 0$.

The primary motivation for considering this problem comes from the question of whether the rich and well-developed theory of stochastic averaging within the setting of Itô SDEs, seen in Chapter 4, extends to the nonlinear filtering framework. More specifically, we try to ascertain whether the stochastic "averaging" of the signal process $\{X_t^\epsilon\}$ (which as a consequence has weak convergence of $\{X_t^\epsilon\}$ to a limiting diffusion $\{\bar{X}_t\}$) entails weak convergence of the corresponding nonlinear filters $\{\pi_t^\epsilon\}$ to some well-defined $\mathcal{P}(\mathbb{R}^d)$-valued limit.

To get some idea of possible limits of the process $\{\pi_t^\epsilon,\ t \in [0,T]\}$, we suppose that the conditions postulated for Theorem 4.4.72 are in force. It then follows that (i) there is a weak solution $\{(\bar{\Omega}, \bar{\mathcal{F}}, \{\bar{\mathcal{F}}_t\}, \bar{P}), (\bar{X}_t, \hat{W}_t)\}$ of the SDE

$$d\bar{X}_t^i = b^i(\bar{X}_t)\,dt + \sum_{j=1}^{r} B^{ij}(\bar{X}_t)\,d\bar{W}_t^j + \sum_{j=1}^{d} c^{ij}(\bar{X}_t)\,d\bar{V}_t^j, \quad 1 \le i \le d, \quad t \in [0,T], \tag{5.1.5}$$

with

$$\bar{X}_0 = x_0 \qquad \text{a.s.} \tag{5.1.6}$$

where $\{\hat{W}_t\} \equiv \{(\bar{W}_t, \bar{V}_t),\ t \in [0,T]\}$ is an $\mathbb{R}^{r+d}$-valued $\{\bar{\mathcal{F}}_t\}$-Wiener process , (ii) the process $\{X^\epsilon,\ t \in [0,T]\}$, given by (5.1.1), converges in law to the limit $\{\bar{X}_t,\ t \in [0,T]\}$ as $\epsilon \to 0$, and (iii) this SDE has the property of uniqueness in law (this results from Condition 4.1.61). The Wiener process $\{\bar{V}_t\}$ in (5.1.5) has arisen in the course of averaging over the fast process $\{Z_t\}$, while $\{\bar{W}_t\}$ may be regarded as a natural counterpart of $\{W_t\}$ in (5.1.1). Using this latter Wiener process we define an $\mathbb{R}^r$-dimensional observation process

$$\bar{Y}_t := \bar{W}_t + \int_0^t h(\bar{X}_u)\,du, \quad t \in [0,T], \tag{5.1.7}$$

and an observation filtration by

$$\mathcal{F}_t^{\bar{Y}} := \sigma\{\bar{Y}_s,\ 0 \le s \le t\} \vee \mathcal{N}(\bar{P}).$$

By Lemma 2.3.5 there is a $\mathcal{P}(\mathbb{R}^d)$-valued, corlol, and $\{\mathcal{F}^{\bar{Y}}_{t+}\}$-optional process $\{\bar{\pi}_t,\ t \in [0,T]\}$ such that

$$\bar{\pi}_t f = \mathrm{E}[f(\bar{X}_t)|\mathcal{F}^{\bar{Y}}_{t+}] \quad a.s. \tag{5.1.8}$$

for each $t \in [0,T]$ and $f \in B(\mathbb{R}^d)$; we regard this process as the *nonlinear filter* for the signal $\{\bar{X}_t,\ t \in [0,T]\}$ corresponding to the observation process $\{\bar{Y}_t,\ t \in [0,T]\}$. The structure of (5.1.1), (5.1.2), and (5.1.5) suggests that $\{\pi^\epsilon_t,\ t \in [0,T]\}$ may converge in law to the nonlinear filter $\{\bar{\pi}_t,\ t \in [0,T]\}$ as $\epsilon \to 0$, and our goal in the present chapter is to show that this is indeed the case. In addition to the conditions imposed in connection with Theorem 4.4.72, we shall postulate the further conditions:

Condition 5.1.73. The mapping $h : \mathbb{R}^d \to \mathbb{R}^r$ is continuous and uniformly bounded.

Condition 5.1.74. The mapping $c : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ in the factorization (4.4.39) (see also the SDE (5.1.5)) is continuous, and $\hat{a}(x) = c(x)c^T(x)$ is strictly positive definite for each $x \in \mathbb{R}^d$.

Remark 5.1.75. From Condition 4.4.71 and Remark 4.1.60 we see that there is a constant $C \in [0,\infty)$ such that

$$\max_{i,j,k}\left\{|b^i(x)|, |B^{ij}(x)|, |c^{ik}(x)|\right\} \leq C[1+|x|], \qquad \forall\, x \in \mathbb{R}^d.$$

We impose Conditions 5.1.73 and 5.1.74 to avail ourselves of the uniqueness results for the nonlinear filter equations established in Chapter 3 (see Theorem 3.3.34 and the associated Conditions 3.3.31, 3.3.32 and 3.3.33), which will be essential for this chapter. The postulate in Condition 5.1.73 that $h(\cdot)$ be *continuous* is not in fact needed for Theorem 3.3.34, which requires that $h(\cdot)$ be only Borel-measurable, but will play a role in several of the technical developments in the present chapter.

Remark 5.1.76. To improve readability in the rest of this chapter we will collectively refer to the conditions needed for Theorem 4.4.72, together with Conditions

5.1.73 and 5.1.74, as Condition **AII**. That is, Condition **AII** encompasses all of the Conditions 4.1.50, 4.1.51, 4.1.52, 4.1.54, 4.1.55, 4.1.61, 4.4.71, 5.1.73, and 5.1.74.

We are now able to state the main result of this chapter:

Theorem 5.1.77. *Suppose that Condition* **AII** *holds. Then*

$$\lim_{\epsilon \to 0+} \mathcal{L}((\pi^\epsilon, Y^\epsilon)) = \mathcal{L}((\bar{\pi}, \bar{Y})) \qquad \textit{in} \qquad \mathcal{P}(C_{\mathcal{P}(\mathbb{R}^d)\times\mathbb{R}^r}[0,T]). \tag{5.1.9}$$

5.2 Preliminaries

In this section we introduce some preliminary considerations before beginning the task of establishing Theorem 5.1.77.

Define $\mathbb{R}^r$-valued processes $\{I_t^\epsilon\}$, $\epsilon \in (0,1]$, and $\{\bar{I}_t\}$, by

$$I_t^\epsilon := Y_t^\epsilon - \int_0^t \pi_u^\epsilon h\, du, \quad t \in [0,T], \tag{5.2.10a}$$

$$\bar{I}_t := \bar{Y}_t - \int_0^t \bar{\pi}_u h\, du, \quad t \in [0,T]. \tag{5.2.10b}$$

Remark 5.2.78. These are innovations processes, and it follows from Lemma 2.4.11 that $\{(I_t^\epsilon, \mathcal{F}_{t+}^{Y^\epsilon})\}$ and $\{(\bar{I}_t, \mathcal{F}_{t+}^{\bar{Y}})\}$ are $\mathbb{R}^r$-valued Wiener processes on $(\Omega, \mathcal{F}, P)$ and $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$ respectively. In view of path continuity it follows that $\{(I_t^\epsilon, \mathcal{F}_t^{Y^\epsilon})\}$ and $\{(\bar{I}_t, \mathcal{F}_t^{\bar{Y}})\}$ are $\mathbb{R}^r$-valued Wiener processes on $(\Omega, \mathcal{F}, P)$ and $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$ respectively.

Remark 5.2.79. Since the mapping

$$(\nu_\cdot, V_\cdot) \in C_{\mathcal{P}(\mathbb{R}^d)\times\mathbb{R}^r}[0,T] \longmapsto \left(\nu_\cdot, V_\cdot + \int_0^\cdot \nu_s h\, ds\right) \in C_{\mathcal{P}(\mathbb{R}^d)\times\mathbb{R}^r}[0,T],$$

is continuous (see Fact C.1.109), we see from (5.2.10) and the continuous mapping theorem (e.g. Corollary 3.1.9 of Ethier and Kurtz [8]), that Theorem 5.1.77 follows when we have proved

Theorem 5.2.80. *Suppose that Condition* **AII** *holds. Then*

$$\lim_{\epsilon\to 0+} \mathcal{L}((\pi^\epsilon, I^\epsilon)) = \mathcal{L}((\bar\pi, \bar I)) \qquad in \qquad \mathcal{P}(C_{\mathcal{P}(\mathbb{R}^d)\times\mathbb{R}^r}[0,T]). \tag{5.2.11}$$

Thus, we must establish Theorem 5.2.80. To this end we first develop the normalized filter equation for the nonlinear filter $\{\bar\pi_t\}$ given by (5.1.8). From (5.1.5), (4.4.39), and Itô's formula, we see that the process

$$\bar M_t^f := \phi(\bar X_t) - \int_0^t \mathcal{L}\phi(\bar X_s)\,ds, \quad t\in[0,T],$$

is an $\{\bar{\mathcal{F}}_t\}$-martingale for each $\phi \in C_c^\infty(\mathbb{R}^d)$, where $\mathcal{L}$ is defined by (4.1.12), (4.1.11), and (4.1.10). Furthermore, by Remark 2.4.9, for

$$\mathcal{B}_k\phi(x) := \sum_{j=1}^d B^{jk}(x)\partial_j\phi(x), \quad \forall x\in\mathbb{R}^d,\ \phi\in C_c^\infty(\mathbb{R}^d),\ 1\le k\le r \tag{5.2.12}$$

we have

$$\langle \bar M^f, \bar W^k\rangle_t = \int_0^t \mathcal{B}_i\phi(X_s)\,ds, \quad 1\le k\le r, \quad t\in[0,T]. \tag{5.2.13}$$

Therefore, Theorem 2.4.9 shows that, for each $\phi\in C_c^\infty(\mathbb{R}^d)$, we have

$$\bar\pi_t\phi = \bar\pi_0\phi + \int_0^t \bar\pi_s(\mathcal{L}\phi)\,ds + \int_0^t \sum_{k=1}^r [\bar\pi_s(h^k\phi + \mathcal{B}_k\phi) - (\bar\pi_s h^k)(\bar\pi_s\phi)]\,d\bar I_s^k, \quad t\in[0,T]. \tag{5.2.14}$$

We next introduce a $\mathcal{P}(\mathbb{R}^d)\times S$-valued process $\{\mu_t^\epsilon\}$, which is auxiliary to the nonlinear filter $\{\pi_t^\epsilon\}$ (recall that S is the compact metric space in which the fast perturbation process $\{Z_t^\epsilon\}$ in (5.1.1) takes values — see Conditions 4.1.51 and 4.1.52), and which will be needed in the proof of Theorem 5.2.80. The $\mathbb{R}^d\times S$-valued process $\{(X_t^\epsilon, Z_t^\epsilon)\}$ is corlol (recall Remark 4.1.53), hence we can use Lemma 2.3.5 to find a $\mathcal{P}(\mathbb{R}^d\times S)$-valued, corlol, and $\{\mathcal{F}_{t+}^{Y^\epsilon}\}$-optional process $\{\mu_t^\epsilon,\ t\in[0,T]\}$ on $(\Omega,\mathcal{F},P)$ such that

$$\mu_t^\epsilon g = \mathrm{E}[g(X_t^\epsilon, Z_t^\epsilon)|\mathcal{F}_{t+}^{Y^\epsilon}] \quad a.s. \quad \forall t\in[0,T],\ g\in B(\mathbb{R}^d\times S). \tag{5.2.15}$$

In view of (5.1.4) and (5.2.15), we see that the nonlinear filter $\{\pi_t^\epsilon\}$ and the process $\{\mu_t^\epsilon\}$ are related by

$$\pi_t^\epsilon f = \mu_t^\epsilon(f \otimes 1) \quad \text{a.s.} \quad \forall t \in [0,T],\ f \in B(\mathbb{R}^d). \tag{5.2.16}$$

We next establish a normalized filter equation for $\{\mu_t^\epsilon\}$. Recall from Proposition 4.2.67 that for every $f \in \tilde{\mathcal{D}}$ (see (4.2.19)) the process

$$M_t^{\epsilon,f} := f(X_t^\epsilon, Z_t^\epsilon) - \int_0^t \mathcal{C}^\epsilon f(X_s^\epsilon, Z_s^\epsilon)\, ds, \quad t \in [0,T]$$

is an $\{\mathcal{F}_t^{W,Z^\epsilon}\}$-martingale. Define the operators $\mathcal{E}_k$, $k = 1, 2, \ldots, r$, on $\tilde{\mathcal{D}}$ by

$$\mathcal{E}_k f(x,z) := \sum_{j=1}^d B^{jk}(x)\partial_j f(x,z), \quad (x,z) \in \mathbb{R}^d \times S,\ f \in \tilde{\mathcal{D}},\ 1 \le k \le r. \tag{5.2.17}$$

Dependence of $\{M_t^{\epsilon,f}\}$ on $\{W_t\}$ is characterized in the next result, whose (rather technical) proof is placed in Appendix C.

Lemma 5.2.81. *Suppose Condition* **AII** *(see Remark 5.1.76) holds. Then for every $\epsilon \in (0,1]$ and $f \in \tilde{\mathcal{D}}$ (see (4.2.19)), we have*

$$\langle M_t^{\epsilon,f}, W^k\rangle_t = \int_0^t \mathcal{E}_k f(X_s^\epsilon, Z_s^\epsilon)\, ds, \quad 0 \le t \le T,\ 1 \le k \le r.$$

From (5.2.10a) and (5.2.16) we have

$$I_t^\epsilon = Y_t^\epsilon - \int_0^t \mu_u^\epsilon(h \otimes 1)\, du, \quad t \in [0,T]. \tag{5.2.18}$$

Now Proposition 4.2.67, Lemma 5.2.81, Theorem 2.4.12, and (5.2.18) show that $\{\mu_t^\epsilon\}$ satisfies the following normalized filter equation: $\forall f \in \tilde{\mathcal{D}}$, $\epsilon \in (0,1]$ we have

$$\mu_t^\epsilon f = \mu_0^\epsilon f + \int_0^t \mu_s^\epsilon(\mathcal{C}^\epsilon f)\, ds + \int_0^t \sum_{k=1}^r [\mu_s^\epsilon((h^k \otimes 1)f + \mathcal{E}_k f) - (\mu_s^\epsilon(h^k \otimes 1))(\mu_s^\epsilon f)]\, d(I^\epsilon)_s^k, \tag{5.2.19}$$

for every $t \in [0,T]$. A result which will soon be needed is the following (the proof is in Appendix C):

Lemma 5.2.82. *Suppose that Condition* **AII** *holds (see Remark 5.1.76). The* $\mathcal{P}(\mathbb{R}^d)$*-valued processes* $\{\pi_t^\epsilon\}$ *and* $\{\bar{\pi}_t\}$ *(see* (5.1.4) *and* (5.1.8)*) are continuous and* $\{\mathcal{F}_t^{Y^\epsilon}\}$ *and* $\{\mathcal{F}_t^{\bar{Y}}\}$*-adapted respectively.*

We will also need to associate martingale problems with the normalized filter equations (5.2.14) and (5.2.19). To allow sufficient flexibility for later applications, we choose a general structure of the filter equations, formalized in the following Definition 5.2.84, which includes both (5.2.14) and (5.2.19) as special cases:

Condition 5.2.83. *Let* E *be a metric space, and let* $c^k \in \bar{C}(E)$, $1 \le k \le r$, *be fixed. Also let* $\mathcal{G}, \mathcal{H}_k \subset \bar{C}(E) \times \bar{C}(E)$, $1 \le k \le r$, *be operators with a common domain* $\mathcal{D}(\mathcal{G}, \mathcal{H}) \subset \bar{C}(E)$.

The following generalizes the notion of weak solution in Definition 3.2.19:

Definition 5.2.84. *Suppose Condition 5.2.83. The pair* $\{(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P), (\nu_t, V_t)\}$ *is a weak solution of the normalized filter equation corresponding to* $(\mathcal{G}, \mathcal{H}, c)$, *if the following holds:*

1. $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ *is a complete filtered probability space;*
2. $\{V_t,\ t \in [0,T]\}$ *is an* $\mathbb{R}^r$*-valued* $\{\mathcal{F}_t\}$*-Wiener process on* $(\Omega, \mathcal{F}, P)$*;*
3. $\{\nu_t,\ t \in [0,T]\}$ *is a* $\mathcal{P}(E)$*-valued, corlol,* $\{\mathcal{F}_t\}$*-adapted process, such that for every* $\phi \in \mathcal{D}(\mathcal{G}, \mathcal{H})$,

$$\nu_t\phi = \nu_0\phi + \int_0^t \nu_s(\mathcal{G}\phi)\,ds + \sum_{k=1}^r \int_0^t R_{\mathcal{H}_k}(\phi, c^k, \nu_s)\,dV_s^k, \quad t \in [0,T], \tag{5.2.20}$$

where

$$R_{\mathcal{H}_k}(\phi, c^k, \nu) := \nu(c^k\phi + \mathcal{H}_k\phi) - (\nu c^k)(\nu\phi), \quad \forall \nu \in \mathcal{P}(E),\ \forall \phi \in \mathcal{D}(\mathcal{G}, \mathcal{H}). \tag{5.2.21}$$

We next write down a martingale problem for the $\mathcal{P}(E)$-valued process $\{\nu_t\}$ in Definition 5.2.84. Our approach is motivated by Hijab [12] (although a similar idea had been used before in studying the Fleming-Viot process of population genetics — see Dawson [6] and the references therein). Define

$$\mathcal{D}(\mathbb{H}(\mathcal{G},\mathcal{H},c)) := \{\Phi \in \bar{C}(\mathcal{P}(E)) : \Phi(\nu) = H(\nu\phi_1, \nu\phi_2, \dots, \nu\phi_n),\ \forall \nu \in \mathcal{P}(E),$$
$$\text{for some } n \in \mathbb{N},\ (\phi_i)_{i=1}^n \subset \mathcal{D}(\mathcal{G},\mathcal{H}),\ H \in C_c^\infty(\mathbb{R}^n)\}, \tag{5.2.22}$$

and for each $\Phi \in \mathcal{D}(\mathbb{H}(\mathcal{G},\mathcal{H},c))$ define (recall (5.2.21))

$$\mathbb{H}(\mathcal{G},\mathcal{H},c)(\Phi)(\nu) := \sum_{i=1}^n \partial_i H(\nu\phi_1, \dots, \nu\phi_n)\nu(\mathcal{G}\phi_i)$$
$$+ \frac{1}{2}\sum_{k=1}^r \sum_{i,j=1}^n \partial_i\partial_j H(\nu\phi_1, \dots, \nu\phi_n) R_{\mathcal{H}_k}(\phi_i, c^k, \nu) R_{\mathcal{H}_k}(\phi_j, c^k, \nu), \quad \forall \nu \in \mathcal{P}(E). \tag{5.2.23}$$

Since $c^k \in \bar{C}(E)$, $k = 1, 2, \dots, r$, it follows at once that

$$\mathbb{H}(\mathcal{G},\mathcal{H},c) \subset \bar{C}(\mathcal{P}(E)) \times \bar{C}(\mathcal{P}(E)). \tag{5.2.24}$$

The next result shows that the $\mathcal{P}(E)$-valued process $\{\nu_t\}$ solves the martingale problem for the operator $\mathbb{H}(\mathcal{G},\mathcal{H},c)$ (see Definition D.2.134). The proof, which is given in Appendix C.1, is just an easy consequence of Itô's formula and (5.2.20):

Lemma 5.2.85. *Suppose Condition 5.2.83, and let the pair $\{(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P), (\nu_t, V_t)\}$ be a weak solution of the normalized filter equation corresponding to $(\mathcal{G},\mathcal{H},c)$ (see Definition 5.2.84). Then the process*

$$M_t^\Phi := \Phi(\nu_t) - \int_0^t \mathbb{H}(\mathcal{G},\mathcal{H},c)(\Phi)(\nu_u)\,du, \quad t \in [0,T]$$

is an $\{\mathcal{F}_t\}$-martingale for each $\Phi \in \mathcal{D}(\mathbb{H}(\mathcal{G},\mathcal{H},c))$.

Useful properties of $\mathcal{D}(\mathbb{H}(\mathcal{G},\mathcal{H},c))$ are summarized in the following two results, the proofs of which are placed in Appendix C.1.

Lemma 5.2.86. *Suppose that Condition 5.2.83 holds. Then $\mathcal{D}(\mathbb{H}(\mathcal{G},\mathcal{H},c))$ is an algebra in $\bar{C}(\mathcal{P}(E))$ that includes constant functions. Moreover, if $\mathcal{D}(\mathcal{G},\mathcal{H})$ is separating (see Definition D.1.122), then $\mathcal{D}(\mathbb{H}(\mathcal{G},\mathcal{H},c))$ separates points in $\mathcal{P}(E)$ (see Definition D.1.121), and if $\mathcal{D}(\mathcal{G},\mathcal{H})$ is convergence determining (see Definition D.1.126) then $\mathcal{D}(\mathbb{H}(\mathcal{G},\mathcal{H},c))$ strongly separates points in $\mathcal{P}(E)$ (see Definition D.1.125).*

Corollary 5.2.87. *Suppose that E is a compact metric space and $\mathcal{D}(\mathcal{G},\mathcal{H})$ is separating in Condition 5.2.83. Then $\mathcal{D}(\mathbb{H}(\mathcal{G},\mathcal{H},c))$ is dense in $\bar{C}(\mathcal{P}(E))$.*

To prove Theorem 5.2.80 we need a martingale problem not just for the process $\{\nu_t\}$, but rather for the pair $\{(\nu_t, V_t)\}$ in Definition 5.2.84. Again following Hijab [12], put

$$\mathcal{D}(\hat{\mathbb{H}}(\mathcal{G},\mathcal{H},c)) := \operatorname{span}\{\Phi\otimes g : \Phi\in\mathcal{D}(\mathbb{H}(\mathcal{G},\mathcal{H},c)),\ g\in C_c^\infty(\mathbb{R}^r)\cup\{1\}\} \tag{5.2.25a}$$

$$\begin{aligned}\hat{\mathbb{H}}(\mathcal{G},\mathcal{H},c)(\Psi)(\nu,y) :=& g(y)\mathbb{H}(\mathcal{G},\mathcal{H},c)(\Phi)(\nu)\\ &+\sum_{i=1}^n\sum_{k=1}^r \partial_i H(\nu\phi_1,\dots,\nu\phi_n)R_{\mathcal{H}_k}(\phi_i,c^k,\nu)\partial_k g(y)+\frac{1}{2}\Phi(\nu)\Delta g(y),\\ &\Psi := \Phi\otimes g\in\mathcal{D}(\hat{\mathbb{H}}(\mathcal{G},\mathcal{H},c)),\ (\nu,y)\in\mathcal{P}(E)\times\mathbb{R}^r.\end{aligned} \tag{5.2.25b}$$

Remark 5.2.88. From (5.2.24) it follows that

$$\hat{\mathbb{H}}(\mathcal{G},\mathcal{H},c)\subset\bar{C}(\mathcal{P}(E)\times\mathbb{R}^r)\times\bar{C}(\mathcal{P}(E)\times\mathbb{R}^r).$$

The following result, which is an easy consequence of Lemma 5.2.85 and Itô's product formula (see Appendix C.1 for the proof), shows that the $\mathcal{P}(E)\times\mathbb{R}^r$-valued process $\{(\nu_t, V_t)\}$ solves the martingale problem for the operator $\hat{\mathbb{H}}(\mathcal{G},\mathcal{H},c)$:

Lemma 5.2.89. *Suppose Condition 5.2.83 holds, and the pair* $\{(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P), (\nu_t, V_t)\}$ *is a weak solution of the normalized filter equation corresponding to* $(\mathcal{G}, \mathcal{H}, c)$ *(see Definition 5.2.84). Then the process*

$$\hat{M}_t^{\Psi} := \Psi(\nu_t, V_t) - \int_0^t \hat{\mathbb{H}}(\mathcal{G}, \mathcal{H}, c)(\Psi)(\nu_u, V_u)\, du, \qquad t \in [0, T]$$

is an $\{\mathcal{F}_t\}$*-martingale for each* $\Psi \in \mathcal{D}(\hat{\mathbb{H}}(\mathcal{G}, \mathcal{H}, c))$.

Remark 5.2.90. (i) Take $E \equiv \mathbb{R}^d$, $c^k \equiv h^k$ and $\mathcal{H}_k \equiv \mathcal{B}_k$, $k = 1, 2, \ldots, r$, $\mathcal{G} \equiv \mathcal{L}$, and $\mathcal{D}(\mathcal{G}, \mathcal{H}) \equiv C_c^\infty(\mathbb{R}^d)$ in Condition 5.2.83. From Remark 5.2.78, Lemma 5.2.82, and (5.2.14), we see that $\{(\bar{\Omega}, \bar{\mathcal{F}}, \{\mathcal{F}_t^{\bar{Y}}\}, \bar{P}), (\bar{\pi}_t, \bar{I}_t)\}$ is a weak solution of the normalized filter equation corresponding to $(\mathcal{L}, \mathcal{B}, h)$ (in the sense of Definition 5.2.84). Thus Lemma 5.2.89 shows that the $\mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^r$-valued process $\{(\bar{\pi}_t, \bar{I}_t),\ t \in [0, T]\}$ is a solution of the martingale problem for $\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h)$ (see Definition D.2.134). From (5.1.6) and (5.1.8) we see that $\bar{\pi}_0 = \delta_{x_0}$, and, from (5.2.10b), that $\bar{I}_0 = 0$. Thus, if $\mu \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^r)$ denotes the Dirac measure concentrated at the point $(\delta_{x_0}, 0) \in \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^r$, then we have $\mathcal{L}(\bar{\pi}_0, \bar{I}_0) = \mu$, hence $\{(\bar{\pi}_t, \bar{I}_t),\ t \in [0, T]\}$ is a solution of the martingale problem for $(\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h), \mu)$.

(ii) Likewise, with $E \equiv \mathbb{R}^d \times S$, $c^k \equiv h^k \otimes 1$ and $\mathcal{H}_k \equiv \mathcal{E}_k$, $k = 1, 2, \ldots, r$, $\mathcal{G} \equiv \mathcal{C}^\epsilon$, and $\mathcal{D}(\mathcal{G}, \mathcal{H}) \equiv \tilde{\mathcal{D}}$ in Condition 5.2.83, we see from Remark 5.2.78 and (5.2.19) that $\{(\Omega, \mathcal{F}, \{\mathcal{F}_{t+}^{Y^\epsilon}\}, P), (\mu_t^\epsilon, I_t^\epsilon)\}$ is a weak solution of the normalized filter equation corresponding to $(\mathcal{C}^\epsilon, \mathcal{E}, h \otimes 1)$. Thus Lemma 5.2.89 shows that the $\mathcal{P}(\mathbb{R}^d \times S) \times \mathbb{R}^r$-valued process $\{(\mu_t^\epsilon, I_t^\epsilon),\ t \in [0, T]\}$ is a solution of the martingale problem for $\hat{\mathbb{H}}(\mathcal{C}^\epsilon, \mathcal{E}, h \otimes 1)$.

5.3 A Convergence Theorem of Bhatt and Karandikar

With the preliminaries of Section 5.2 in place we are now ready to discuss the main result which will be used to establish Theorem 5.2.80, namely:

Theorem 5.3.91 (Theorem 2.1 & Remark 2.2 of Bhatt and Karandikar [3]). *Suppose that E is a complete separable metric space, and $\mathcal{A}$ is a linear operator with domain $\mathcal{D}(\mathcal{A})$ having the following properties:*

(O) $\mathcal{A} \subset \bar{C}(E) \times \bar{C}(E)$.

(I) There exists a countable set $\mathfrak{C} \subset \mathcal{D}(\mathcal{A})$ such that $\{(f, \mathcal{A}f) : f \in \mathcal{D}(\mathcal{A})\}$ is a subset of the bp-closure of $\{(f, \mathcal{A}f) : f \in \mathfrak{C}\}$.

(II) $\mathcal{D}(\mathcal{A})$ is an algebra that separates points in E (see Definition D.1.121) and vanishes nowhere, and there is a countable subset of $\mathcal{D}(\mathcal{A})$ that strongly separates points in E (see Definition D.1.125).

(III) The martingale problem for $\mathcal{A}$ is well-posed (see Definition D.2.134 and Definition D.2.137).

Suppose further that, for some $\mu \in \mathcal{P}(E)$, we have

(IV) the martingale problem for $(\mathcal{A}, \mu)$ has a solution $\{X_t,\ t \in [0,T]\}$ with corlol paths.

(V) there is a sequence $\{X_n(t),\ t \in [0,T]\}$, $n = 1, 2, \ldots$ of E-valued processes with corlol paths such that $\{\mathcal{L}(X_n(t)),\ n = 1, 2, \ldots\}$ is a tight sequence in $\mathcal{P}(E)$ for each $t \in [0,T]$, and $\lim_n \mathcal{L}(X_n(0)) = \mu$ in $\mathcal{P}(E)$.

(VI) for each $f \in \mathcal{D}(\mathcal{A})$, there exist $\mathbb{R}$-valued progressively measurable processes $\{(U_n(t), \mathcal{F}_t^n),\ t \in [0,T]\}$ and $\{(V_n(t), \mathcal{F}_t^n),\ t \in [0,T]\}$, $n = 1, 2, \ldots$, such that

$$U_n(t) - \int_0^t V_n(s)\, ds, \qquad t \in [0,T], \qquad \text{is an } \{\mathcal{F}_t^n\} - \text{martingale}, \tag{5.3.26}$$

$$\lim_n \mathrm{E}\left[\sup_{t\in[0,T]} |U_n(t) - f(X_n(t))|\right] = 0, \tag{5.3.27}$$

$$\sup_n \mathrm{E}\left[\left\{\int_0^T |V_n(s)|^p\,ds\right\}^{1/p}\right] < \infty, \qquad \textit{for some} \quad p \in (1,\infty), \tag{5.3.28}$$

$$\lim_n \mathrm{E}[|\mathcal{A}f(X_n(t)) - V_n(t)|] = 0, \qquad \textit{for each} \quad t \in [0,T]. \tag{5.3.29}$$

Then $\{X_n(t),\ t \in [0,T]\}$ *converges weakly to* $\{X(t),\ t \in [0,T]\}$ *as* $n \to \infty$.

Remark 5.3.92. Comparing Theorem 5.3.91 to the somewhat similar Theorem 4.8.10 of Ethier and Kurtz [8], one observes that the former result postulates only relative compactness of the *one dimensional marginals* of an E-valued sequence of corlol processes $\{X_n\}$, whereas the latter theorem makes the significantly stronger hypothesis of relative compactness of the set of probability measures in $D_E[0,\infty)$ of the whole process X_n. In this respect, Theorem 5.3.91 is considerably easier to use than Theorem 4.8.10 of [8]. This is a (subtle) consequence of the stronger uniqueness hypothesized in Theorem 5.3.91, namely uniqueness of solutions of the martingale problem for $\mathcal{A}$ within the class of *progressively measurable* processes (see (III) of Theorem 5.3.91), whereas Theorem 4.8.10 of [8] postulates only uniqueness within the class of *corlol* candidate solutions. Although the uniqueness hypothesis in Theorem 5.3.91 is more difficult to verify than the uniqueness hypthesized by Theorem 4.8.10 of [8], it is nevertheless the case that Theorem 5.3.91 is much better suited to establishing weak convergence of nonlinear filters because its hypothesis on relative compactness are so much easier to verify.

Remark 5.3.93. In order to establish Theorem 5.2.80 it is enough to show that

$$\lim_{n\to\infty} \mathcal{L}((\pi^{\epsilon_n}, I^{\epsilon_n})) = \mathcal{L}((\bar{\pi}, \bar{I})) \qquad \text{in} \qquad \mathcal{P}(C_{\mathcal{P}(\mathbb{R}^d)\times\mathbb{R}^r}[0,T]). \tag{5.3.30}$$

for an arbitrary sequence $\{\epsilon_n\} \subset (0,1]$ such that $\lim_n \epsilon_n = 0$. We henceforth regard this sequence as fixed, and, to lighten the notation, we denote $\{\pi_t^{\epsilon_n}\}$ and $\{I_t^{\epsilon_n}\}$ by $\{\pi_t^n\}$ and $\{I_t^n\}$ respectively.

Remark 5.3.94. We are going to use Theorem 5.3.91 to show (5.3.30). To this end, we will identify E and $(\mathcal{A}, \mathcal{D}(\mathcal{A}))$ in Theorem 5.3.91 with the complete separable metric space $\mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^r$ and linear operator $(\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h), \mathcal{D}(\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h)))$ respectively, and will verify (0), (I), (II) and (III) of Theorem 5.3.91 for the operator $(\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h), \mathcal{D}(\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h)))$. We will further identify the probability measure μ in Theorem 5.3.91 with the Dirac measure in $\mathcal{P}(\mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^r)$ concentrated at the point $(\delta_{x_0}, 0)$, the process $\{X_t,\ t \in [0,T]\}$ in Theorem 5.3.91(IV) with $\{(\bar{\pi}_t, \bar{I}_t),\ t \in [0,T]\}$, and the process $\{X_n(t),\ t \in [0,T]\}$ in Theorem 5.3.91(V) with $\{(\pi_t^n, I_t^n),\ t \in [0,T]\}$, and will then verify the remaining conditions (IV), (V) and (VI).

Verification of (0) in Theorem 5.3.91: Since $h^k \in \bar{C}(\mathbb{R}^d)$, $k = 1, 2, \ldots, r$ (see Condition 5.1.73), by Remark 5.2.88 we have

$$\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h) \subset \bar{C}(\mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^r) \times \bar{C}(\mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^r),$$

as required.

Verification of (I) in Theorem 5.3.91: This is an immediate consequence of the following simple result, the proof of which is located in Appendix C.2:

Lemma 5.3.95. *The set*

$$\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h) \subset \bar{C}(\mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^r) \times \bar{C}(\mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^r)$$

is separable (in the supremum norm of $\bar{C}(\mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^r) \times \bar{C}(\mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^r)$*).*

Verification of (II) in Theorem 5.3.91: By Lemma 5.2.86 we know that $\mathcal{D}(\mathbb{H}(\mathcal{L}, \mathcal{B}, h))$ is an algebra that includes constant functions, hence vanishes nowhere. Since $\mathcal{D}(\mathcal{L}, \mathcal{B}) := C_c^\infty(\mathbb{R}^d)$ is separating (see Fact D.1.129) we see from Lemma 5.2.86 that $\mathcal{D}(\mathbb{H}(\mathcal{L}, \mathcal{B}, h))$ separates points in $\mathcal{P}(\mathbb{R}^d)$. Then it follows from (5.2.25a) that $\mathcal{D}(\hat{\mathbb{H}}(\mathcal{L},\mathcal{B},h))$ is an algebra that vanishes nowhere and separates points in $\mathcal{P}(\mathbb{R}^d)\times\mathbb{R}^r$. It remains to see that $\mathcal{D}(\hat{\mathbb{H}}(\mathcal{L},\mathcal{B},h))$ includes a countable subset that strongly separates points in $\mathcal{P}(\mathbb{R}^d)\times\mathbb{R}^r$. Since $\mathcal{D}(\mathcal{L},\mathcal{B}) := C_c^\infty(\mathbb{R}^d)$ is convergence determining (see Fact D.1.129), we see from Lemma 5.2.86 that $\mathcal{D}(\mathbb{H}(\mathcal{L},\mathcal{B},h))$ strongly separates points in $\mathcal{P}(\mathbb{R}^d)$, and since $\mathcal{D}(\mathbb{H}(\mathcal{L},\mathcal{B},h))$ is separable (see Lemma 5.3.95) it follows that $\mathcal{D}(\mathbb{H}(\mathcal{L},\mathcal{B},h))$ includes a countable subset which strongly separates points in $\mathcal{P}(\mathbb{R}^d)$. Since $\hat{C}(\mathbb{R}^r)$ is separable in the supremum norm, we see that $C_c^\infty(\mathbb{R}^r)$ is likewise separable, and $C_c^\infty(\mathbb{R}^r)$ is easily seen to strongly separate points in $\mathbb{R}^r$. Hence $C_c^\infty(\mathbb{R}^r)$ includes a countable subset that strongly separates points in $\mathbb{R}^r$. Now it follows from (5.2.25a) that $\mathcal{D}(\hat{\mathbb{H}}(\mathcal{L},\mathcal{B},h))$ includes a countable subset that strongly separates points in $\mathcal{P}(\mathbb{R}^d)\times\mathbb{R}^r$.

Verification of (III) in Theorem 5.3.91: This is the most important of the conditions associated with Theorem 5.3.91, and is verified by the following result, which in turn relies in an essential way on the uniqueness Theorem 3.3.34(ii) giving uniqueness in joint law for the normalized filter equation. The proof of Theorem 5.3.96 is in Appendix C.3.

Theorem 5.3.96. *Suppose Condition* **AII** *(see Remark 5.1.76). Then the martingale problem for $\hat{\mathbb{H}}(\mathcal{L},\mathcal{B},h)$ is well-posed.*

Verification of (IV) in Theorem 5.3.91: As noted in Remark 5.3.94, we take $\mu\in\mathcal{P}(\mathcal{P}(\mathbb{R}^d)\times\mathbb{R}^r)$ in Theorem 5.3.91 to be the Dirac measure concentrated at the point $(\delta_{x_0},0)$. Then, from Remark 5.2.90(i), we see that $\{(\bar{\pi}_t,\bar{I}_t),\ t\in[0,T]\}$ is a solution of the martingale problem for $(\hat{\mathbb{H}}(\mathcal{L},\mathcal{B},h),\mu)$. Moreover, the paths of $\{(\bar{\pi}_t,\bar{I}_t),\ t\in[0,T]\}$ are necessarily continuous (by Lemma 5.2.82 and (5.2.10b)), thus corlol.

Verification of (V) in Theorem 5.3.91: As noted in Remark 5.3.94, we identify $\{X_n(t),\ t\in[0,T]\}$ in Theorem 5.3.91(V) with $\{(\pi_t^n,I_t^n),\ t\in[0,T]\}$, which, in view of Lemma 5.2.82 and (5.2.10a), is continuous, hence corlol. From (5.1.1), (5.1.4), and (5.2.10a), it follows that $\mathcal{L}(\pi_0^n, I_0^n) = \mu$, $\forall\ n = 1, 2, \ldots$. Now, verification of (V) is completed by the following result, which is established in Appendix C.2:

Lemma 5.3.97. *Suppose Condition* **AII** *(see Remark 5.1.76). Then* $\{\mathcal{L}(\pi_t^n),\ n = 1, 2, \ldots\}$ *is tight in* $\mathcal{P}(\mathcal{P}(\mathbb{R}^d))$ *for each* $t \in [0, T]$.

Verification of (VI) in Theorem 5.3.91:

Remark 5.3.98. Here we will use the perturbed test functions $f^{\epsilon,\phi}$ furnished by Proposition 4.2.68, and the $\mathcal{P}(\mathbb{R}^d \times S)$-valued process $\{\mu_t^\epsilon,\ t \in [0, T]\}$ given by (5.2.15). In keeping with Remark 5.3.30, we lighten the notation and put $\mathcal{C}^n$ for $\mathcal{C}^{\epsilon_n}$ (see (4.2.20)), $f^{n,\phi}$ for $f^{\epsilon_n,\phi}$ (see (4.2.22)), $\{\mu_t^n\}$ for $\{\mu_t^{\epsilon_n}\}$, and $\mathcal{F}_t^n$ for $\mathcal{F}_{t+}^{Y^{\epsilon_n}}$ (see (5.1.3)).

Fix arbitrary $\Psi \in \mathcal{D}(\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h))$ (see (5.2.25a)) of the form

$$\Psi(\nu, y) = (\Phi \otimes g)(\nu, y), \quad \forall (\nu, y) \in \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^r \tag{5.3.31}$$

for some

$$\Phi \in \mathcal{D}(\mathbb{H}(\mathcal{L}, \mathcal{B}, h)), \quad g \in \operatorname{span}\{1, C_c^\infty(\mathbb{R}^r)\},$$

with Φ given by

$$\Phi(\nu) = H(\nu\phi_1, \ldots, \nu\phi_m), \quad \nu \in \mathcal{P}(\mathbb{R}^d), \tag{5.3.32}$$

for some positive integer m, some $H \in C_c^\infty(\mathbb{R}^m)$, $\phi_i \in C_c^\infty(\mathbb{R}^d)$, $1 \le i \le m$ (see (5.2.22)). For each $n \in \mathbb{N}$ define

$$\Phi_n(\mu) := H(\mu f^{n,\phi_1}, \ldots, \mu f^{n,\phi_m}), \quad \mu \in \mathcal{P}(\mathbb{R}^d \times S), \tag{5.3.33}$$

where $f^{n,\phi_i} \in \tilde{\mathcal{D}}$ is given by Proposition 4.2.68 with $\epsilon := \epsilon_n$ and $\phi := \phi_i$. Since

$$\mathcal{C}^n, \mathcal{E}_k \subset \bar{C}(\mathbb{R}^d \times S) \times \bar{C}(\mathbb{R}^d \times S), \quad \forall n \in \mathbb{N},\ 1 \le k \le r,$$

(see (5.2.17)) and $f^{n,\phi_i} \in \tilde{\mathcal{D}}$, it follows that

$$\Phi_n \in \mathcal{D}\big(\mathbb{H}(\mathcal{C}^n, \mathcal{E}, h \otimes 1)\big), \quad \forall n \in \mathbb{N}, \tag{5.3.34}$$

hence

$$\Phi_n \otimes g \in \mathcal{D}\big(\hat{\mathbb{H}}(\mathcal{C}^n, \mathcal{E}, h \otimes 1)\big), \quad \forall n \in \mathbb{N}. \tag{5.3.35}$$

For each $t \in [0, T]$, $n \in \mathbb{N}$, put

$$U_n(t) := (\Phi_n \otimes g)(\mu_t^n, I_t^n), \tag{5.3.36a}$$

$$V_n(t) := \hat{\mathbb{H}}(\mathcal{C}^n, \mathcal{E}, h \otimes 1)(\Phi_n \otimes g)(\mu_t^n, I_t^n). \tag{5.3.36b}$$

From Remark 5.2.90(ii), we know that the $\mathcal{P}(\mathbb{R}^d \times S) \times \mathbb{R}^r$-valued $\{\mathcal{F}_t^n\}$-progressively measurable process $\{(\mu_t^n, I_t^n),\ t \in [0, T]\}$ is a solution of the martingale problem for $\hat{\mathbb{H}}(\mathcal{C}^n, \mathcal{E}, h \otimes 1)$, hence (5.3.35), (5.3.36a), and (5.3.36b) show that

$$U_n(t) - \int_0^t V_n(s)\, ds, \quad t \in [0, T]$$

is a $\{\mathcal{F}_t^n\}$-martingale, as required for (5.3.26). To verify the remaining conditions (5.3.27), (5.3.28), and (5.3.29) we need the following elementary fact, the proof of which is given in Appendix C.2:

Fact 5.3.99. *(a) There is a constant $C \in [0, \infty)$ such that*

$$\sup_{\substack{t \in [0,T] \\ \omega \in \Omega}} |\Phi_n(\mu_t^n) - \Phi(\pi_t^n)| \leq C\epsilon_n, \quad \forall n \in \mathbb{N}. \tag{5.3.37}$$

(b) $\sup_{n,\omega,t} |V_n(t)| < \infty$.

(c) There is a constant $C \in [0, \infty)$ such that

$$\sup_{n,\omega} |V_n(t) - \hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h)(\Psi)(\pi_t^n, I_t^n)| < C\,\epsilon_n, \quad \forall t \in [0, T]. \tag{5.3.38}$$

To verify (5.3.27), we see from (5.3.36a) and (5.3.31) that

$$U_n(t) - \Psi(\pi_t^n, I_t^n) = [\Phi_n(\mu_t^n) - \Phi(\pi_t^n)]g(I_t^n), \quad \forall t \in [0,T],\ n \in \mathbb{N}. \tag{5.3.39}$$

and thus, by uniform boundedness of $g(\cdot)$, (5.3.37) and (5.3.39),

$$\lim_{n\to\infty} \mathrm{E}\left[\sup_{0\le t\le T} |U_n(t) - \Psi(\pi_t^n, I_t^n)|\right] = 0,$$

as required. Now (5.3.28) is an immediate consequence of Fact 5.3.99(b) and (5.3.29) follows from Fact 5.3.99(c) and the dominated convergence theorem.

All conditions of Theorem 5.3.91 have now been checked, and hence we conclude that

$$\lim_{n\to\infty} \mathcal{L}((\pi^n, I^n)) = \mathcal{L}((\bar{\pi}, \bar{I})) \qquad \text{in} \qquad \mathcal{P}(D_{\mathcal{P}(\mathbb{R}^d)\times\mathbb{R}^r}[0,T]). \tag{5.3.40}$$

But $\{(\pi_t^n, I_t^n),\ t \in [0,T]\}$ and $\{(\bar{\pi}_t, \bar{I}_t),\ t \in [0,T]\}$ are *continuous* $\mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^r$-valued processes (see Lemma 5.2.82, (5.2.10a), and (5.2.10b)) thus we get (5.3.30) from (5.3.40) and Corollary 3.3.2 of Ethier and Kurtz [8]. We have thus established Theorem 5.2.80 (see Remark 5.3.93), and therefore, in view of Remark 5.2.79, we have also proved Theorem 5.1.77.

Remark 5.3.100. Kurtz and Ocone [22] introduced the notion of a *filtered martingale problem*, which has subsequently been used by a number of authors for establishing uniqueness for the SDEs of nonlinear filtering (see Remark 3.3.36). The main feature of the filtered martingale problem is that it uniquely characterizes the joint law of the pair (optimal filter, observation) in the nonlinear filtering framework, that is, every solution of the filtered martingale problem with the same initial distribution as the pair (optimal filter, observation) has necessarily the same finite-dimensional distributions as (optimal filter, observation). However, in order for a process $\{(\mu_t, U_t)\}$ to be a solution of the filtered martingale problem, the measure-valued component $\{\mu_t\}$ has to be adapted to the filtration generated by the process $\{U_t\}$. Since this typically does not hold when the pair $\{(\mu_t, U_t)\}$ arises as a weak limit of a sequence of processes, the filtered martingale problem is not well suited for establishing weak convergence of nonlinear filters. For this reason we cannot use the filtered martingale problem as a tool for proving Theorem 5.1.77, and instead use the martingale problem formalized by Lemma 5.2.89, which is closer to classical martingale problem ideas in that it behaves well with respect to weak convergence.

We end this chapter with some remarks in which we compare the main convergence result of this chapter (Theorem 5.1.77) with other works on convergence of nonlinear filters in the established literature.

Remark 5.3.101. Bhatt, Kallianpur and Karandikar [1] study a problem which comprises the following main ingredients:

(i) for each $n = 1, 2, \dots$, the signal $\{X_t^n\}$ is a corlol process on $(\Omega_n, \mathcal{F}_n, P_n)$, taking values in a complete separable metric space E_1.

(ii) for each $n = 1, 2, \dots$, the corresponding observation $\{Y_t^n\}$ is an $\mathbb{R}^r$-valued process on $(\Omega_n, \mathcal{F}_n, P_n)$ given by

$$Y_t^n := W_t^n + \int_0^t h_n(X_s^n)\, ds, \tag{5.3.41}$$

for some $\mathbb{R}^r$-valued Wiener process $\{W_t^n\}$ which is *independent* of $\{X_t^n\}$, and some continuous sensor function $h_n : E_1 \to \mathbb{R}^r$ which satisfies the finite-energy condition

$$\mathrm{E}\left[\int_0^T |h_n(X_s^n)|^2\, ds\right] < \infty.$$

(iii) $\{\bar{X}_t\}$ is an E_1-valued corlol signal on $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$, with corresponding $\mathbb{R}^r$-valued observation $\{\bar{Y}_t\}$ given by

$$\bar{Y}_t := \bar{W}_t + \int_0^t h(\bar{X}_s)\, ds, \tag{5.3.42}$$

for some $\mathbb{R}^r$-valued Wiener process $\{\bar{W}_t\}$ which is *independent* of $\{\bar{X}_t\}$, and some continuous sensor function $h : E_1 \to \mathbb{R}^r$ which satisfies the finite-energy condition

$$\mathrm{E}\left[\int_0^T |h(\bar{X}_s)|^2\, ds\right] < \infty.$$

Let $\{\pi_t^n\}$ be the $\mathcal{P}(E_1)$-valued nonlinear filter of the signal $\{X_t^n\}$ corresponding to the observation process $\{Y_t^n\}$ (recall Remark 2.3.7) and, likewise, let $\{\bar{\pi}_t\}$ be the $\mathcal{P}(E_1)$-valued nonlinear filter of the signal $\{\bar{X}_t\}$ corresponding to the observation process $\{\bar{Y}_t\}$. The following is a special case of Theorem 9.4(ii) from [1]:

Theorem 5.3.102. *Suppose that E_1 is locally compact, $\{X_t^n\}$ solves the martingale problem for some operator $\mathcal{A}_n \subset \bar{C}(E_1) \times \bar{C}(E_1)$, $\{\bar{X}_t\}$ solves the martingale problem for some operator $\mathcal{A} \subset \bar{C}(E_1) \times \bar{C}(E_1)$, and the operators $\mathcal{A}$, $\mathcal{A}_n$, have common domain $\mathcal{D} \subset \bar{C}(E_1)$ for all $n = 1, 2, \ldots$. If*

(a) for each $f \in \mathcal{D}$ one has $\sup_n \|\mathcal{A}_n f\| < \infty$;

(b) $\{X_t^n\}$ converges weakly to $\{\bar{X}_t\}$;

(c) h_n converges to h uniformly on compact subsets of E_1;

(d) the martingale problem for $\mathcal{A}$ is well-posed,

then $\{\pi_t^n\}$ converges weakly to $\{\bar{\pi}_t\}$.

This is a very general convergence result, but nevertheless fails to include the the problem that we address in this chapter for the following reason:

The Wiener process $\{W_t^n\}$ is assumed to be independent of the signal $\{X_t^n\}$, $n = 1, 2, \ldots$, and the Wiener process $\{\bar{W}_t\}$ is assumed to be independent of the signal $\{\bar{X}_t\}$. In fact, this assumption is the key element in the proof of Theorem 5.3.102, since it permits direct computation of the nonlinear filters $\{\pi_t^n\}$ and $\{\bar{\pi}_t\}$ by means of the Kallianpur-Striebel formula, and the weak convergence of the nonlinear filters is then a direct consequence of these computations. Now fix some $\{\epsilon_n\} \subset (0, 1]$ with

$\lim_n \epsilon_n = 0$, identify the probability spaces $(\Omega_n, \mathcal{F}_n, P_n)$ in Theorem 5.3.102 with the common probability space $(\Omega, \mathcal{F}, P)$ on which $X^n := X^{\epsilon_n}$ and $Y^n := Y^{\epsilon_n}$ in (5.1.1) and (5.1.2) are defined, and identify $\{W_t^n\}$ in (5.3.41) with the common Wiener process $\{W_t\}$ in (5.1.2). Then we see that the independence necessary for Theorem 5.3.102 fails to hold because of the third term on the right-hand side of (5.1.1) and the second term on the right-hand side of (5.1.5). Moreover, the natural counterpart of the operators $\mathcal{A}_n$ in Theorem 5.3.102 are the operators $\mathfrak{C}^\epsilon$, $\epsilon := \epsilon_n$, in (4.2.20). It is evident that the right side of (4.2.20) goes to infinity with $n \to \infty$, so that we can never verify the uniform bound in Theorem 5.3.102(a).

Remark 5.3.103. Goggin [11] studies the following problem: Given complete separable metric spaces E_1 and E_2, together with Borel-measurable mappings $(X, Y) : (\Omega, \mathcal{F}, P) \to E_1 \times E_2$ and $(X^n, Y^n) : (\Omega_n, \mathcal{F}_n, P_n) \to E_1 \times E_2$ such that $\{(X^n, Y^n)\}$ converges weakly to (X, Y), determine general conditions which ensure that $\{\mathrm{E}^n[f(X^n)|Y^n]\}$ converges weakly to $\mathrm{E}[f(X)|Y]$ for each $f \in \bar{C}(E_1)$. We first note that weak convergence of (X^n, Y^n) to (X, Y) is by itself *not enough* to ensure weak convergence of the corresponding conditional expectations, as shown in Goggin [11], and that additional conditions are definitely needed. Such conditions are provided by Theorem 2.1 of [11], and essentially postulate that there must exist some $Q_n \equiv P_n$ on $(\Omega_n, \mathcal{F}_n)$, and some $Q \equiv P$ on $(\Omega, \mathcal{F})$, such that X^n and Y^n are Q_n-independent, and X and Y are Q-independent. Essentially, Theorem 2.1 of Goggin [11] says that if the Q_n-distribution of $(X^n, Y^n, dQ_n/dP_n)$ converges weakly to the Q-distribution of $(X, Y, dQ/dP)$, then the P_n distribution of $\mathrm{E}^n[f(X^n)|Y^n]$ converges weakly to the P-distribution of $\mathrm{E}[f(X)|Y]$ for each $f \in \bar{C}(E_1)$. This result has obvious applicability to the convergence of nonlinear filters, and indeed has been used to this end by Goggin ([11], Section 4). However, there are two basic limitations on using this theorem for the convergence problem of this chapter, namely: (i) it appears necessary to assume independence of the Wiener process

$\{W_t\}$ in (5.1.2) and the signal $\{X_t^\epsilon\}$ in (5.1.1) (by taking $B \equiv 0$ in (5.1.1)) for it is only under this condition that we can construct a probability measure $Q^\epsilon \equiv P$ (by a Girsanov transformation) which renders $\{X_t^\epsilon\}$ and $\{Y_t^\epsilon\}$ independent; (ii) even if one adopts the restriction in (i), the very structure of the result entails that it yields weak convergence of π_t^ϵ to $\bar{\pi}_t$ *for each fixed* t (i.e. weak convergence of the single-dimensional marginals of $\{\pi_t^\epsilon\}$ to the corresponding single-dimensional marginals of $\{\bar{\pi}_t\}$) rather than convergence of the whole nonlinear filter process $\{\pi_t^\epsilon\}$ to the limit process $\{\bar{\pi}_t\}$, as desired.

Remark 5.3.104. Kushner [25] looks at the singularly perturbed SDEs

$$\begin{aligned} dX_t^\epsilon &= G(X_t^\epsilon, Z_t^\epsilon)\,dt + \sigma_1(X_t^\epsilon, Z_t^\epsilon)\,dV_t^1 \\ dZ_t^\epsilon &= \epsilon^{-1}H(X_t^\epsilon, Z_t^\epsilon)\,dt + \epsilon^{-1/2}\sigma_2(X_t^\epsilon, Z_t^\epsilon)\,dV_t^2, \end{aligned}$$

where $\{V_t^1\}$ and $\{V_t^2\}$ are independent Wiener processes, and the coefficients are sufficiently well behaved so that there exists an $\mathbb{R}^{d+D}$-valued weak solution $\{(X_t^\epsilon, Z_t^\epsilon)\}$ that is unique in law for each $\epsilon \in (0,1)$. The process $\{X_t^\epsilon\}$ is regarded as the signal, and an $\mathbb{R}^r$-valued observation process $\{Y_t^\epsilon\}$ is defined by

$$Y_t^\epsilon = W_t + \int_0^t h(X_s^\epsilon, Z_s^\epsilon)\,ds,$$

where $\{W_t\}$ is Wiener process, and $\{(V_t^1, V_t^2)\}$ and $\{W_t\}$ are independent. Thus, in particular, $\{(X_t^\epsilon, Z_t^\epsilon)\}$ and the observation Wiener process $\{W_t\}$ are independent. Let $\{\pi_t^\epsilon\}$ be the nonlinear filter for the signal $\{X_t^\epsilon\}$ given the observation $\{Y_t^\epsilon\}$, namely $\{\pi_t^\epsilon\}$ is a corlol $\{\mathcal{F}_{t+}^{Y^\epsilon}\}$-optional process such that

$$\pi_t^\epsilon \phi = \mathrm{E}[\phi(X_t^\epsilon)|\mathcal{F}_{t+}^{Y^\epsilon}] \ \ a.s.$$

for each $\phi \in \bar{C}(\mathbb{R}^d)$. Under sufficiently strong hypothesis (see Theorem 1 of Kushner [25]) it is shown that $\{(X_t^\epsilon, Y_t^\epsilon)\}$ converges weakly to the $\mathbb{R}^{d+r}$-valued process

$\{(\bar{X}_t, \bar{Y}_t)\}$ that satisfies

$$d\bar{X}_t = \bar{G}(\bar{X}_t)\,dt + \bar{\sigma}(\bar{X}_t)\,d\bar{V}_t$$
$$\bar{Y}_t = \bar{W}_t + \int_0^t \bar{h}(\bar{X}_s)\,ds,$$

for appropriately defined functions $\bar{G}$, $\bar{\sigma}$, and $\bar{h}$ (the exact form of which is unimportant for our discussion), and where $\{\bar{V}_t\}$ and $\{\bar{W}_t\}$ are *independent* Wiener processes. Let $\{\bar{\pi}_t\}$ be the nonlinear filter of the signal $\{\bar{X}_t\}$ given the observation $\{\bar{Y}_t\}$. Since $\{\bar{W}_t\}$ and $\{\bar{X}_t\}$ are independent, for each $t \in [0, T]$ there exists a Borel-measurable mapping $\bar{F}_t : C_{\mathbb{R}^r}[0, t] \to C_{\mathcal{P}(\mathbb{R}^d)}[0, t]$ such that

$$\bar{\pi}_t = \bar{F}_t(\bar{Y}_\cdot) \;\; a.s.$$

The mapping $\bar{F}_t$ actually has a closed-form expression and is just the Kallianpur-Striebel formula. Now define a $\mathcal{P}(\mathbb{R}^d)$-valued "hybrid" filter

$$\bar{\pi}_t^\epsilon := \bar{F}_t(Y_\cdot^\epsilon), \quad \forall t \in [0, T].$$

We emphasize that $\{\bar{\pi}_t^\epsilon\}$ is not the nonlinear filter $\{\pi_t^\epsilon\}$ introduced earlier, but is a hybrid entity obtained by using the statistics of the limiting process $\{\bar{X}_t\}$ which effectively determines the form of the Kallianpur-Striebel formula $\bar{F}_t$, together with the "non-limiting" observation process $\{Y_t^\epsilon\}$. In Theorem 2 of Kushner [25] it is shown that $\{(\bar{\pi}_t^\epsilon, Y_t^\epsilon)\}$ converges weakly to $\{(\bar{\pi}_t, \bar{Y}_t)\}$. This analysis depends crucially on the availability of the Kallianpur-Striebel formula for $\bar{F}_t$, and this in turn requires the postulated independence of $\{\bar{W}_t\}$ and $\{\bar{X}_t\}$. On the other hand, our main result (Theorem 5.1.77) is concerned with convergence of the *actual* filter $\{\pi_t^\epsilon\}$, rather than the hybrid filter $\{\bar{\pi}_t^\epsilon\}$, and in a setting where there is dependence of $\{\bar{X}_t\}$ on the Wiener process $\{\bar{W}_t\}$, so that the Kallianpur-Striebel formula does not apply.

Remark 5.3.105. In a work related to [26], Kushner [27] also looks at convergence of nonlinear filters when *wide-band approximations* are used for the Wiener processes occurring in the dynamics of the signal $\{X_t^\epsilon\}$ and the observation process $\{Y_t^\epsilon\}$. More specifically, the signal is modeled by the ODE

$$\dot{X}_t^\epsilon = G(X_t^\epsilon) + B(X_t^\epsilon)\,\dot{V}_t^\epsilon,$$

where

$$V_t^\epsilon := \frac{1}{\epsilon}\int_0^t \psi(u/\epsilon^2)\,du,$$

for some uniformly bounded, stationary, zero-mean, and uniform-mixing process $\{\psi_t\}$, and the observation $\{Y_t^\epsilon\}$ is given by

$$Y_t^\epsilon = \dot{W}_t^\epsilon + h(X_t^\epsilon),$$

where

$$W_t^\epsilon := \frac{1}{\epsilon}\int_0^t \xi(u/\epsilon^2)\,du,$$

for some uniformly bounded, stationary, zero-mean, and uniform-mixing process $\{\xi_t\}$. To simplify the notation we will henceforth suppose that X_t^ϵ, Y_t^ϵ, W_t^ϵ and V_t^ϵ are *real* valued. Correlation between the signal $\{X_t^\epsilon\}$ and observation $\{Y_t^\epsilon\}$ is introduced by supposing that

$$\int_{-\infty}^{\infty} \mathrm{E}\begin{bmatrix}\psi(t)\\ \xi(t)\end{bmatrix}\begin{bmatrix}\psi(0) & \xi(0)\end{bmatrix} dt = \begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}$$

for some number ρ. Under suitable conditions (see A.1, A.2, A.3, A.4, and A.5 in Kushner [27]) it may be shown that $\{(V_t^\epsilon, W_t^\epsilon, X_t^\epsilon, Y_t^\epsilon)\}$ converges weakly to $\{(\bar{V}_t, \bar{W}_t, \bar{X}_t, \bar{Y}_t)\}$, where $\{\bar{V}_t\}$ and $\{\bar{W}_t\}$ are correlated Wiener processes with $\mathrm{E}[V_1 W_1] = \rho$, and

$$d\bar{X}_t = \bar{G}(X_t) + \sigma(X_t)\,d\bar{V}_t,$$
$$\bar{Y}_t = \bar{W}_t + \int_0^t h(\bar{X}_s)\,ds,$$

for some appropriately defined function G. An approximate "wide-band" nonlinear filter $\{\bar{\pi}_t^\epsilon\}$ is defined as follows: Let $\{\bar{\sigma}_t^\epsilon\}$ be the $\mathcal{M}^+(\mathbb{R})$-valued process which solves the family of "measure-valued ODE's"

$$\frac{d}{dt}(\bar{\sigma}_t^\epsilon\phi) = \bar{\sigma}_t^\epsilon(\mathcal{A}\phi) + \left(\bar{\sigma}_t^\epsilon(\sigma\rho\,\partial\phi)\right)(\dot{Y}_t^\epsilon) - \frac{1}{2}\bar{\sigma}_t^\epsilon\left(\rho^2\sigma\,\partial(\sigma\,\partial\phi) + \rho\sigma\,\partial(\phi h) + \phi h^2\right), \tag{5.3.43}$$

parametrized by $\phi \in \mathcal{D}(\mathcal{A})$, where $\mathcal{A}$ is the second-order differential operator for the diffusion $\{\bar{X}_t\}$, and let $\{\bar{\pi}_t^\epsilon\}$ be the $\mathcal{P}(\mathbb{R})$-valued process given by

$$\bar{\pi}_t^\epsilon := \frac{\bar{\sigma}_t^\epsilon}{\bar{\sigma}_t^\epsilon 1}. \tag{5.3.44}$$

Like the hybrid nonlinear filter in Remark 5.3.104, the process $\{\bar{\pi}_t^\epsilon\}$ given by (5.3.43) and (5.3.44) is also a hybrid, since (5.3.43) uses both the generator $\mathcal{A}$ which characterizes the *limit* process $\{\bar{X}_t\}$ and the non-limiting observation process $\{Y_t^\epsilon\}$, which appears as the "exogenous" term that drives (5.3.43). In Theorem 4 of Kushner [27] it is shown, subject to fairly stringent hypothesis on G, σ, and h, that $\{(\bar{\pi}_t^\epsilon, Y_t^\epsilon)\}$ converges weakly to $\{(\bar{\pi}_t, \bar{Y}_t)\}$, where, as usual, $\{\bar{\pi}_t\}$ is the nonlinear filter of the limiting signal $\{\bar{X}_t\}$ given the observation $\{\bar{Y}_t\}$. Contrasting this result with our Theorem 5.1.77, we see that it deals with a different convergence issue, namely convergence of the approximate filter $\{\bar{\pi}_t^\epsilon\}$ determined by (5.3.43) and (5.3.44), whereas we study convergence of the *actual* filter $\{\pi_t^\epsilon\}$ given $\pi_t^\epsilon\phi = \mu_t^\epsilon(\phi \otimes 1)$, where $\{\mu_t^\epsilon\}$ is the process determined by (5.2.19). Notice that the generator in (5.2.19) is $\mathcal{C}^\epsilon$, which becomes unbounded as $\epsilon \to 0$ (see (4.2.20)), unlike the corresponding generator $\mathcal{A}$ in (5.3.43) which is fixed. This feature makes the asymptotic analysis a good deal more delicate, and calls for the use of powerful convergence results such as Theorem 5.3.91.

Appendix A

Appendices for Chapter 3

A.1 Proofs of Technical Results

Proof of Fact 3.4.40: Fix some $\alpha \in (1, \infty)$. Since $\tilde{\sigma}_0$ takes values in $\mathcal{P}(\mathbb{R}^d)$, we see from (3.2.19) with $\phi \equiv 1$ that

$$\tilde{\sigma}_t 1 = 1 + \int_0^t \sum_{k=1}^r \left(\frac{\tilde{\sigma}_s h^k}{\tilde{\sigma}_s 1} \right) (\tilde{\sigma}_s 1)\, d\tilde{Y}_s^k, \quad t \in [0, T].$$

This gives (see Exercise IV.3.10(1) of Revuz and Yor [30])

$$\tilde{\sigma}_t 1 = \mathcal{E}\left(\sum_{k=1}^r \left(\frac{\tilde{\sigma} h^k}{\tilde{\sigma} 1} \bullet \tilde{Y}^k \right) \right)_t, \quad t \in [0, T], \tag{A.1.1}$$

and hence

$$\begin{aligned} |\tilde{\sigma}_t 1|^\alpha &= \tilde{M}_t \exp\left(\frac{\alpha(\alpha-1)}{2} \sum_{k=1}^r \int_0^t \left| \frac{\tilde{\sigma}_s h^k}{\tilde{\sigma}_s 1} \right|^2 ds \right) \\ &\leq \tilde{M}_t \exp\left(\frac{\alpha(\alpha-1)}{2} \|h\| T \right), \quad \forall\, t \in [0, T], \end{aligned} \tag{A.1.2}$$

for

$$\tilde{M}_t := \mathcal{E}\left(\alpha \sum_{k=1}^r \left(\frac{\tilde{\sigma} h^k}{\tilde{\sigma} 1} \bullet \tilde{Y}^k \right) \right)_t, \quad t \in [0, T].$$

Now Condition 3.3.33 ensures that the processes $\{(\tilde{\sigma}_t h^k)/(\tilde{\sigma}_t 1),\ t \in [0,T]\}$ are uniformly bounded (by $\|h^k\|$), and therefore $\{(\tilde{M}_t, \tilde{\mathcal{F}}_t),\ t \in [0,T]\}$ is a continuous martingale on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{Q})$, with $\tilde{M}_0 = 1$. Taking $\tilde{Q}$-expectations in (A.1.2) then gives

$$\mathrm{E}^{\tilde{Q}}[|\tilde{\sigma}_t 1|^{\alpha}] \leq \exp\left(\frac{\alpha(\alpha-1)}{2}\|h\|T\right), \quad \forall\, t \in [0,T]. \tag{A.1.3}$$

Again, by (A.1.1) and uniform-boundedness of the processes $\{(\tilde{\sigma}_t h^k)/(\tilde{\sigma}_t 1),\ t \in [0,T]\}$ we see that $\{(\tilde{\sigma}_t 1, \tilde{\mathcal{F}}_t),\ t \in [0,T]\}$ is a continuous martingale on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{Q})$, which, in light of (A.1.3), is L_α-bounded. Thus, by Doob's inequality, there is some $\gamma(\alpha) \in [0,\infty)$ such that (3.4.29) holds. □

Proof of Lemma 3.5.47: Since $\{\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{Q}), (\tilde{\sigma}^i_t, \tilde{Y}_t)\}$, $i = 1,2$ are weak solutions of the unnormalized filter equation, for some fixed $f_1, f_2 \in C_c^\infty(\mathbb{R}^d)$ we have

$$\tilde{\sigma}^i_t f_i = \tilde{\sigma}^i_0 f_i + \int_0^t \tilde{\sigma}^i_u(\mathcal{A} f_i)\, du + \sum_k \int_0^t \tilde{\sigma}^i_u(h^k f_i + \mathcal{B}_k f_i)\, d\tilde{Y}^k_u, \quad t \in [0,T],\ i = 1,2. \tag{A.1.4}$$

By Itô's formula from (A.1.4) we have

$$(\tilde{\sigma}^1_t f_1)(\tilde{\sigma}^2_t f_2) = (\tilde{\sigma}^1_0 f_1)(\tilde{\sigma}^2_0 f_2) + \int_0^t (\tilde{\sigma}^1_u f_1)\, d(\tilde{\sigma}^2_\cdot f_2)_u + \int_0^t (\tilde{\sigma}^2_u f_2)\, d(\tilde{\sigma}^1_\cdot f_1)_u + \langle \tilde{\sigma}^1 f_1, \tilde{\sigma}^2 f_2 \rangle_t, \tag{A.1.5}$$

hence

$$\begin{aligned}(\tilde{\sigma}^1_t f_1)(\tilde{\sigma}^2_t f_2) = {} & (\tilde{\sigma}^1_0 f_1)(\tilde{\sigma}^2_0 f_2) + \int_0^t (\tilde{\sigma}^1_u f_1)(\tilde{\sigma}^2_u(\mathcal{A} f_2))\, du + \int_0^t (\tilde{\sigma}^2_u f_2)(\tilde{\sigma}^1_u(\mathcal{A} f_1))\, du \\ & + \sum_k \int_0^t \big[(\tilde{\sigma}^1_u(h^k f_1))(\tilde{\sigma}^2_u(h^k f_2)) + (\tilde{\sigma}^1_u(h^k f_1))(\tilde{\sigma}^2_u(\mathcal{B}_k f_2)) \\ & + (\tilde{\sigma}^1_u(\mathcal{B}_k f_1))(\tilde{\sigma}^2_u(h^k f_2)) + (\tilde{\sigma}^1_u(\mathcal{B}_k f_1))(\tilde{\sigma}^2_u(\mathcal{B}_k f_2))\big]\, du + M_t,\end{aligned} \tag{A.1.6}$$

where

$$M_t := \sum_k \int_0^t \left\{(\tilde{\sigma}^1_u f_1)[\tilde{\sigma}^2_u(h^k f_2 + \mathcal{B}_k f_2)] + (\tilde{\sigma}^2_u f_2)[\tilde{\sigma}^1_u(h^k f_1 + \mathcal{B}_k f_1)]\right\} d\tilde{Y}^k_u, \quad t \in [0,T],$$

is a $\{\tilde{\mathcal{F}}_t\}$-continuous local martingale on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{Q})$. Observe by Fact 3.4.40 and Cauchy-Schwarz inequality that

$$\begin{aligned}
\mathrm{E}^{\tilde{Q}}\left[\int_0^t |\tilde{\sigma}_u^1 f_1|^2 |\tilde{\sigma}_u^2(h^k f_2 + \mathcal{B}_k f_2)|^2 \, du\right] \\
&\le \|f_1\|^2 \left[\|h^k f_2\| + \|\mathcal{B}_k f_2\|\right]^2 \mathrm{E}^{\tilde{Q}}\left[\int_0^t |\tilde{\sigma}_u^1 1|^2 |\tilde{\sigma}_u^2 1|^2 \, du\right] \\
&\le \gamma(4) T \|f_1\|^2 \left[\|h^k f_2\| + \|\mathcal{B}_k f_2\|\right]^2 < \infty,
\end{aligned}$$

and so $\{M_t\}$ is a square-integrable $\{\tilde{\mathcal{F}}_t\}$-martingale on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{Q})$, with $M_0 \equiv 0$. Therefore, by taking $\tilde{Q}$-expectations in (A.1.6), we get

$$\begin{aligned}
\mathrm{E}^{\tilde{Q}}[(\tilde{\sigma}_t^1 f_1)(\tilde{\sigma}_t^2 f_2)] = {} & \mathrm{E}^{\tilde{Q}}[(\tilde{\sigma}_0^1 f_1)(\tilde{\sigma}_0^2 f_2)] + \int_0^t \mathrm{E}^{\tilde{Q}}[(\tilde{\sigma}_u^1 f_1)(\tilde{\sigma}_u^2(\mathcal{A} f_2))] \, du \\
& + \int_0^t \mathrm{E}^{\tilde{Q}}[(\tilde{\sigma}_u^1(\mathcal{A} f_1))(\tilde{\sigma}_u^2 f_2)] \, du \\
& + \sum_k \int_0^t \Big[\mathrm{E}^{\tilde{Q}}[(\tilde{\sigma}_u^1(h^k f_1))(\tilde{\sigma}_u^2(h^k f_2))] + \mathrm{E}^{\tilde{Q}}[(\tilde{\sigma}_u^1(h^k f_1))(\tilde{\sigma}_u^2(\mathcal{B}_k f_2))] \\
& + \mathrm{E}^{\tilde{Q}}[(\tilde{\sigma}_u^1(\mathcal{B}_k f_1))(\tilde{\sigma}_u^2(h^k f_2))] + \mathrm{E}^{\tilde{Q}}[(\tilde{\sigma}_u^1(\mathcal{B}_k f_1))(\tilde{\sigma}_u^2(\mathcal{B}_k f_2))]\Big] \, du,
\end{aligned} \tag{A.1.7}$$

and thus, from (3.4.34),

$$\begin{aligned}
\nu_t^{12}(f_1 \otimes f_2) = {} & \nu_0^{12}(f_1 \otimes f_2) + \int_0^t [\nu_u^{12}(f_1 \otimes \mathcal{A} f_2) + \nu_u^{12}(\mathcal{A} f_1 \otimes f_2)] \, du \\
+ \sum_k \int_0^t [& \nu_u^{12}(h^k f_1 \otimes h^k f_2) + \nu_u^{12}(h^k f_1 \otimes \mathcal{B}_k f_2) + \nu_u^{12}(\mathcal{B}_k f_1 \otimes h^k f_2) + \nu_u^{12}(\mathcal{B}_k f_1 \otimes \mathcal{B}_k f_2)] \, du,
\end{aligned}$$

which establishes (3.5.59). The proofs of the corresponding facts for $\nu_\cdot^{11}$ and $\nu_\cdot^{22}$ follow in exactly the same way. □

Proof of Lemma 3.5.48: Fix some $f_1, f_2 \in C_c^\infty(\mathbb{R}^d)$, and let $f := f_1 \otimes f_2$ (recall Remark 3.4.42). Write $x \in \mathbb{R}^{2d}$ in the form $x = (x_1, x_2)$, for $x_1, x_2 \in \mathbb{R}^d$. We then have the first derivatives

$$\partial_i f(x) = f_2(x_2) \partial_i f_1(x_1), \qquad \partial_{i+d} f(x) = f_1(x_1) \partial_i f_2(x_2), \tag{A.1.8}$$

for $i = 1, \dots, d$, and the second derivatives

$$\left.\begin{aligned} \partial_i\partial_j f(x) &= f_2(x_2)\partial_i\partial_j f_1(x_1), \\ \partial_{i+d}\partial_{j+d} f(x) &= f_1(x_1)\partial_i\partial_j f_2(x_2), \\ \partial_i\partial_{j+d} f(x) &= \partial_i f_1(x_1)\partial_j f_2(x_2), \end{aligned}\right\} \tag{A.1.9}$$

for $i, j = 1, \dots, d$. From (3.5.58) we can put

$$\tilde{A}(f_1 \otimes f_2)(x) = I_1(x) + I_2(x) + \bar{h}(x)f(x), \tag{A.1.10}$$

where $I_1(x)$ is the sum of all terms of $\tilde{A}(f_1 \otimes f_2)(x)$ which involve second-order derivatives of the function f (see (3.2.8a) and (3.2.8b)) namely

$$\begin{aligned} I_1(x) := &\frac{1}{2}\sum_{i,j=1}^{d}[BB^T + cc^T]^{ij}(x_2)f_1(x_1)\partial_i\partial_j f_2(x_2) \\ &+ \frac{1}{2}\sum_{i,j=1}^{d}[BB^T + cc^T]^{ij}(x_1)f_2(x_2)\partial_i\partial_j f_1(x_1) \\ &+ \sum_{k=1}^{r}\left(\sum_{i=1}^{d} B^{ik}(x_1)\partial_i f_1(x_1)\right)\left(\sum_{j=1}^{d} B^{jk}(x_2)\partial_j f_2(x_2)\right), \end{aligned} \tag{A.1.11}$$

and $I_2(x)$ is the sum of all terms of $\tilde{A}(f_1 \otimes f_2)(x)$ which involve first-order derivatives of f namely

$$\begin{aligned} I_2(x) := &\sum_{i=1}^{d} b^i(x_1)f_2(x_2)\partial_i f_1(x_1) + \sum_{i=1}^{d} b^i(x_2)f_1(x_1)\partial_i f_2(x_2) \\ &+ \sum_{k=1}^{r} h^k(x_1)f_1(x_1)\sum_{i=1}^{d} B^{ik}(x_2)\partial_i f_2(x_2) \\ &+ \sum_{k=1}^{r} h^k(x_2)f_2(x_2)\sum_{i=1}^{d} B^{ik}(x_1)\partial_i f_1(x_1). \end{aligned} \tag{A.1.12}$$

For each $x \in \mathbb{R}^{2d}$ put

$$\bar{B}(x) := \begin{bmatrix} B(x_1) \\ B(x_2) \end{bmatrix} \begin{bmatrix} B^T(x_1) & B^T(x_2) \end{bmatrix}, \quad \bar{C}(x) := \begin{bmatrix} cc^T(x_1) & 0 \\ 0 & cc^T(x_2) \end{bmatrix}, \tag{A.1.13}$$

and observe that

$$[\bar{B}]^{ij}(x) = [BB^T]^{ij}(x_1), \quad [\bar{B}]^{(i+d)(j+d)}(x) = [BB^T]^{ij}(x_2), \tag{A.1.14a}$$

$$[\bar{B}]^{i(j+d)}(x) = [\bar{B}]^{(j+d)i}(x) = \sum_{k=1}^{r} B^{ik}(x_1)B^{jk}(x_2), \tag{A.1.14b}$$

and

$$[\bar{C}]^{i,j}(x) = [cc^T]^{ij}(x_1), \quad [\bar{C}]^{(i+d)(j+d)}(x) = [cc^T]^{ij}(x_2), \tag{A.1.15a}$$

$$[\bar{C}]^{i(j+d)} = [\bar{C}]^{(j+d)i} = 0, \tag{A.1.15b}$$

for all $i, j = 1, \ldots, d$. From (A.1.9), (A.1.11), (A.1.14) and (A.1.15),

$$\begin{aligned}
I_1(x) =& \frac{1}{2}\sum_{i,j=1}^{d}[\bar{B}+\bar{C}]^{(i+d)(j+d)}(x)\partial_{i+d}\partial_{j+d}f(x) + \frac{1}{2}\sum_{i,j=1}^{d}[\bar{B}+\bar{C}]^{ij}(x)\partial_i\partial_j f(x) \\
&+\frac{1}{2}\sum_{i,j=1}^{d}[\bar{B}+\bar{C}]^{i(j+d)}(x)\partial_i\partial_{j+d}f(x) + \frac{1}{2}\sum_{i,j=1}^{d}[\bar{B}+\bar{C}]^{(j+d)i}(x)\partial_{j+d}\partial_i f(x) \\
=&\frac{1}{2}\sum_{i,j=1}^{2d}[\bar{B}+\bar{C}]^{ij}(x)\partial_i\partial_j f(x),
\end{aligned}$$

and, from (3.4.40a) and (A.1.13),

$$\bar{a}(x) = \bar{B}(x) + \bar{C}(x),$$

hence

$$I_1(x) = \frac{1}{2}\sum_{i,j=1}^{2d}\bar{a}^{ij}(x)\partial_i\partial_j f(x). \tag{A.1.16}$$

Similarly, from (A.1.8), (A.1.12), and (3.4.40b),

$$\begin{aligned}
I_2(x) =& \sum_{i=1}^{d}\left[b^i(x_1) + \sum_{k=1}^{r}B^{ik}(x_1)h^k(x_2)\right]\partial_i f(x) \\
&+\sum_{i=1}^{d}\left[b^i(x_2) + \sum_{k=1}^{r}B^{ik}(x_2)h^k(x_1)\right]\partial_{i+d} f(x) \\
=&\sum_{i=1}^{d}\bar{b}^i(x)\partial_i f(x) + \sum_{i=1}^{d}\bar{b}^{i+d}(x)\partial_{i+d}f(x) = \sum_{i=1}^{2d}\bar{b}^i(x)\partial_i f(x),
\end{aligned} \tag{A.1.17}$$

Now (3.5.60) follows from (3.4.41), (A.1.10), (A.1.16) and (A.1.17). □

Proof of Lemma 3.5.49: Fix arbitrary $\epsilon \in (0,\infty)$ and $g \in C_c^\infty(\mathbb{R}^{2d})$. Put

$$B_R := \{x \in \mathbb{R}^{2d} \ : \ |x| \leq R\}, \quad R \in [0,\infty),$$

and fix R such that the support of g is contained in B_R. Also fix some $q \in C_c^\infty(\mathbb{R}^d)$ such that

$$\|q\| \leq 1; \tag{A.1.18a}$$

$$q(z) = 1, \quad \forall\, z \in \mathbb{R}^d, \ \text{ with } |z| \leq R; \tag{A.1.18b}$$

$$q(z) = 0, \quad \forall\, z \in \mathbb{R}^d, \ \text{ with } |z| \geq R\sqrt{2}. \tag{A.1.18c}$$

By Proposition 7.1 in Appendix 7 of Ethier and Kurtz [8], there exists a polynomial $p : \mathbb{R}^{2d} \to \mathbb{R}$ such that

$$\max_{x \in B_{2R}} |g(x) - p(x)| < \epsilon; \tag{A.1.19a}$$

$$\max_{x \in B_{2R}} |\partial_i g(x) - \partial_i p(x)| < \epsilon, \quad \forall\, i = 1, \ldots, 2d; \tag{A.1.19b}$$

$$\max_{x \in B_{2R}} |\partial_i \partial_j g(x) - \partial_i \partial_j p(x)| < \epsilon, \quad \forall\, i,j = 1, \ldots, 2d. \tag{A.1.19c}$$

Since $g(x) = 0$ when $x \notin B_R$, we note from (A.1.19a) that

$$\sup_{x \in B_{2R} \setminus B_R} |p(x)| < \epsilon. \tag{A.1.20}$$

For all $x \in \mathbb{R}^{2d}$, put $x := (x_1, x_2)$, $x_1, x_2 \in \mathbb{R}^d$, and define

$$\bar{q}(x) := q(x_1)q(x_2),$$
$$f(x) := \bar{q}(x)\,p(x).$$

Since $q \in C_c^\infty(\mathbb{R}^d)$ and $p(x)$ is a polynomial in $x = (x_1, x_2)$, it follows that $f \in \bar{\mathcal{D}}$ (recall (3.5.61)). From (A.1.18), we have $\bar{q}(x) = 0$ when $x \notin B_{2R}$, and $\bar{q}(x) = 1$

when $x \in B_R$. Thus

$$\begin{aligned}\|f-g\| &= \sup_{x\in\mathbb{R}^{2d}} |\bar{q}(x)p(x)-g(x)| = \max_{x\in B_{2R}} |\bar{q}(x)p(x)-g(x)| \\ &\le \max_{x\in B_R} |p(x)-g(x)| + \sup_{x\in B_{2R}\setminus B_R} |\bar{q}(x)p(x)-g(x)|,\end{aligned}$$

hence (A.1.19a) and (A.1.20) give

$$\|f-g\| \le \epsilon + \sup_{x\in B_{2R}\setminus B_R} |p(x)| \le 2\epsilon. \tag{A.1.21}$$

Next, consider $\|\bar{A}f-\bar{A}g\|$. From (3.4.41) we have

$$\bar{A}f(x) = \bar{q}(x)\bar{A}p(x) + p(x)\bar{A}\bar{q}(x) + (\nabla p(x))^T\bar{a}(x)\nabla\bar{q}(x). \tag{A.1.22}$$

By the choice of R we have $g(x)=0$ and therefore $\bar{A}g(x)=0$, $\forall x \notin B_R$. Moreover, from (A.1.18c), we have $\bar{q}(x) = 1$, and therefore $\nabla\bar{q}(x) = 0$ and $\bar{A}\bar{q}(x) = 0$, $\forall\ x \in B_R$. Similarly, $\bar{q}(x) = 0$, and therefore $\bar{A}\bar{q}(x) = 0$, $\forall x \notin B_{2R}$. Then, it follows from (A.1.22) that

$$\begin{aligned}\|\bar{A}f-\bar{A}g\| &= \sup_{x\in B_{2R}} |\bar{q}(x)\bar{A}p(x) + p(x)\bar{A}\bar{q}(x) + (\nabla p(x))^T\bar{a}(x)\nabla\bar{q}(x) - \bar{A}g(x)| \\ &\le \sup_{x\in B_R} |\bar{q}(x)\bar{A}p(x) + p(x)\bar{A}\bar{q}(x) + (\nabla p(x))^T\bar{a}(x)\nabla\bar{q}(x) - \bar{A}g(x)| \\ &+ \sup_{x\in B_{2R}\setminus B_R} |\bar{q}(x)\bar{A}p(x) + p(x)\bar{A}\bar{q}(x) + (\nabla p(x))^T\bar{a}(x)\nabla\bar{q}(x)| \\ &= \sup_{x\in B_R} |\bar{A}p(x) - \bar{A}g(x)| \\ &+ \sup_{x\in B_{2R}\setminus B_R} (|\bar{q}(x)\bar{A}p(x)| + |p(x)\bar{A}\bar{q}(x)| + |(\nabla p(x))^T\bar{a}(x)\nabla\bar{q}(x)|).\end{aligned} \tag{A.1.23}$$

Since $\bar{a}$ and $\bar{b}$ are locally bounded, we have

$$C_1 := \sup_{x\in B_{2R}} \left(\sum_{i=1}^{2d} |\bar{b}_i(x)| + \frac{1}{2}\sum_{i,j=1}^{2d} |\bar{a}_{ij}(x)| \right) < \infty.$$

Also let

$$C_2 := \|\bar{q}\| + \sum_i \|\partial_i \bar{q}\| + \sum_{i,j} \|\partial_i \partial_j \bar{q}\| < \infty.$$

Then by (A.1.19)

$$\sup_{x \in B_R} |\bar{A}p(x) - \bar{A}g(x)| < C_1 \epsilon. \tag{A.1.24}$$

Similarly, by (A.1.19) and the fact that $g(x) = 0$, $\forall x \in B_{2R} \setminus B_R$, we obtain

$$\left.\begin{aligned} |\partial_i p(x)| &< \epsilon, \\ |\bar{A}p(x)| &< C_1 \epsilon, \end{aligned}\right\} \forall\, x \in B_{2R} \setminus B_R,$$

and hence, from these bounds and (A.1.20),

$$\sup_{x \in B_{2R} \setminus B_R} (|\bar{q}(x)\bar{A}p(x)| + |p(x)\bar{A}\bar{q}(x)| + |(\nabla p(x))^T \bar{a}(x) \nabla \bar{q}(x)|) \leq \epsilon C_1 + \epsilon C_1 C_2 + \epsilon C_1 C_2. \tag{A.1.25}$$

Now, upon combining (A.1.23), (A.1.24), and (A.1.25) we have

$$\|\bar{A}g - \bar{A}f\| \leq 2\epsilon\,(C_1 + C_1 C_2), \tag{A.1.26}$$

and the result follows. □

A.2 Proof of Theorem 3.4.44

In view of Remark 1 on page 345 of Bhatt and Karandikar [4], with no loss of generality we suppose that $\lambda \in B(\mathbb{R}^q)$ is a nonnegative mapping, and also suppose that the constant K in (3.4.44) is such that $0 \leq \lambda(x) \leq K$, $\forall\, x \in \mathbb{R}^q$.

Remark A.2.106. We use Δ to denote the point at infinity in the one-point compactification $\mathbb{R}^{q*}$ of $\mathbb{R}^q$. Also, members of $B(\mathbb{R}^{q*})$ and operators on $B(\mathbb{R}^{q*})$ will be superscripted with "$*$", and, for $f^* \in B(\mathbb{R}^{q*})$, we write $f^*|_{\mathbb{R}^q}$ to denote the restriction of f^* to the domain $\mathbb{R}^q$.

Define the linear operator $\mathcal{C}^* : \mathcal{D}(\mathcal{C}^*) \subset C(\mathbb{R}^{q*}) \to B(\mathbb{R}^{q*})$ by

$$\mathcal{D}(\mathcal{C}^*) := \{f^* \in C(\mathbb{R}^{q*}) : f^*|_{\mathbb{R}^q} - f^*(\Delta) \in C_c^\infty(\mathbb{R}^q)\}, \tag{A.2.27a}$$

$$\mathcal{C}^* f^*(x) := \mathcal{C}(f^*|_{\mathbb{R}^q} - f^*(\Delta))(x), \qquad \forall x \in \mathbb{R}^q, \quad \forall f^* \in \mathcal{D}(\mathcal{C}^*), \tag{A.2.27b}$$

$$\mathcal{C}^* f^*(\Delta) := 0, \quad \forall f^* \in \mathcal{D}(\mathcal{C}^*). \tag{A.2.27c}$$

Also, define linear operator $\mathcal{G}^* : \mathcal{D}(\mathcal{G}^*) \subset C(\mathbb{R}^{q*}) \to B(\mathbb{R}^{q*})$ by

$$\mathcal{G}^* f^*(x) := \mathcal{C}^* f^*(x) - \lambda(x)(f^*(x) - f^*(\Delta)), \quad \forall\, x \in \mathbb{R}^{q*}, \quad \forall\, f^* \in \mathcal{D}(\mathcal{G}^*) := \mathcal{D}(\mathcal{C}^*). \tag{A.2.28}$$

In order to establish Theorem 3.4.44 it is enough to show that uniqueness holds for the forward equation of the operator $\mathcal{G}^*$ in the following sense: If $\mu^{*,i} : [0,\infty) \to \mathcal{P}(\mathbb{R}^{q*})$, $i = 1, 2$ are such that (i) $\mu_0^{*,1} = \mu_0^{*,2}$; (ii) the mappings $t \mapsto \mu_t^{*,i}(\Gamma)$, $i = 1, 2$, are Borel-measurable on $[0,\infty)$ for each $\Gamma \in \mathcal{B}(\mathbb{R}^{q*})$; and (iii)

$$\mu_t^{*,i} f^* = \mu_0^{*,i} f^* + \int_0^t \mu_s^{*,i}(\mathcal{G}^* f^*)\, ds, \quad \forall t \in [0,\infty), \quad \forall f^* \in \mathcal{D}(\mathcal{G}^*), \quad i = 1, 2, \tag{A.2.29}$$

then $\mu_t^{*,1} = \mu_t^{*,2}$, $\forall\, t \in [0,\infty)$. To see how this gives Theorem 3.4.44, we follow the proof of Theorem 3.4 of Bhatt and Karandikar [4], briefly summarizing their argument for completeness. Let $\mu_\cdot^i : [0,T] \to \mathcal{M}^+(\mathbb{R}^q)$, $i = 1, 2$, be as postulated in Theorem 3.4.44. Without loss of generality suppose that μ_0^1 and μ_0^2 are *probability* measures. Extend the $\mu_\cdot^i$ from $[0,T]$ to $[0,\infty)$ by defining

$$\mu_t^i(\Gamma) := \int_{\mathbb{R}^q} \mathrm{E}_x \left[I_\Gamma(\omega_t) \exp\left(\int_0^{t-T} -\lambda(\omega_s) ds \right) \right] \mu_T^i(dx), \quad t \in [T,\infty), \quad \Gamma \in \mathcal{B}(\mathbb{R}^q),$$

where $P_x \in \mathcal{P}(C_{\mathbb{R}^q}[0,\infty))$, $x \in \mathbb{R}^q$, is the probability measure on the space $C_{\mathbb{R}^q}[0,\infty)$ which solves the martingale problem for $(\mathcal{C}, \delta_x)$, and $\omega_\cdot$ denotes a generic element of $C_{\mathbb{R}^q}[0,\infty)$ (existence and uniqueness of P_x follow from Theorem 8.1.7 of Ethier and Kurtz [8]). Then it is easily checked that

$$\mu_t^i f = \mu_0^i f + \int_0^t \mu_s^i(\mathcal{C}f - \lambda f)\,ds, \quad \forall\, t \in [0,\infty), \quad \forall\, f \in \operatorname{span}\{1, C_c^\infty(\mathbb{R}^q)\}, \quad i = 1,2. \tag{A.2.30}$$

Since $(1,0) \in \mathcal{C}$, we see from (A.2.30) that

$$\mu_t^i(\mathbb{R}^q) = \mu_0^i(\mathbb{R}^q) - \int_0^t \mu_s^i \lambda\, ds, \quad \forall\, t \in [0,\infty),$$

and therefore $\mu_t^i(\mathbb{R}^q) \le 1$, $\forall\, t \in [0,\infty)$ (since μ_0^i is a probability measure and $\lambda(\cdot)$ is nonnegative). Now define $\mu_t^{*,i} \in \mathcal{P}(\mathbb{R}^{q*})$, $t \in [0,\infty)$, $i = 1,2$, as follows: $\mu_t^{*,i}(\Gamma) := \mu_t^i(\Gamma \cap \mathbb{R}^q) + (1 - \mu_t^i(\mathbb{R}^q)) I_\Gamma(\Delta)$, $\forall\, \Gamma \in \mathcal{B}(\mathbb{R}^{q*})$. Then it follows that $\{\mu_t^{*,i}\}$, $i = 1,2$, solve the forward equation (A.2.29), hence the postulated uniqueness for this equation gives $\mu_t^{*,1} = \mu_t^{*,2}$, and so $\mu_t^1 = \mu_t^2$, $\forall\, t \in [0,\infty)$, as required to establish Theorem 3.4.44.

Thus, it remains to show that uniqueness holds for the forward equation of the operator $\mathcal{G}^*$. Define

$$F := \left\{ y = (y_1, y_2, y_3) \in \mathbb{S}_+^{q\times q} \times \mathbb{R}^q \times \mathbb{R} \,:\, |y_1^{ij}| \le K,\ |y_2^i| \le K,\ 0 \le y_3 \le K \right\}.$$

We shall need the following special case of Theorem 2.7(c) from Kurtz [21]:

Theorem A.2.107. *Suppose that $\mathbb{A}^0 : \mathcal{D}(\mathbb{A}^0) \subset \bar{C}(\mathbb{R}^{q*}) \to \bar{C}(\mathbb{R}^{q*} \times F)$ is a linear operator, η is a transition function from $\mathbb{R}^{q*}$ to F, and define*

$$\mathbb{A}_\eta^0 f^*(x) := \int_F \mathbb{A}^0 f^*(x,y)\eta(x,dy), \quad \forall\, x \in \mathbb{R}^{q*}, \quad \forall\, f^* \in \mathcal{D}(\mathbb{A}_\eta^0) := \mathcal{D}(\mathbb{A}^0).$$

Suppose also that (i) $\mathcal{D}(\mathbb{A}^0)$ is closed under multiplication and separates points, (ii) $\mathbb{A}_y^0 f^* \equiv \mathbb{A}^0 f^*(\cdot, y)$ is a pre-generator for each $y \in F$, and (iii) $\mathbb{A}_\eta^0$ satisfies the following separability hypothesis: there exists some countable $\{g_k^*\} \subset \mathcal{D}(\mathbb{A}_\eta^0)$ such that the graph of $\mathbb{A}_\eta^0$ is included within the bp-closure of the linear span of

$\{(g_k^*, \mathbb{A}_\eta^0 g_k^*)\}$. *With these conditions we have the following: if uniqueness holds for the martingale problem for* $\mathbb{A}_\eta^0$ *then uniqueness holds for the forward equation for* $\mathbb{A}_\eta^0$.

We shall use this result to get uniqueness for the forward equation of $\mathcal{G}^*$. Motivated by Example 3.4 of Kurtz and Stockbridge [23], for each $y \in F$ define linear operator $\mathcal{L}_y$ on $\bar{C}(\mathbb{R}^q)$ by

$$\mathcal{L}_y f(x) := \sum_i (1 + |x|)\, y_2^i\, \partial_i f(x) + \frac{1}{2} \sum_{i,j} (1 + |x|^2)\, y_1^{ij}\, \partial_i \partial_j f(x),$$
$$\forall x \in \mathbb{R}^q, \quad f \in \mathcal{D}(\mathcal{L}_y) := C_c^\infty(\mathbb{R}^q). \tag{A.2.31}$$

Also, put $\mathcal{D}(\mathbb{A}^0) := \mathcal{D}(\mathcal{C}^*)$ (see (A.2.27a)) and

$$\mathbb{A}^0 f^*(x,y) := \mathcal{L}_y(f^*|_{\mathbb{R}^q} - f^*(\Delta))(x) - y_3(f^*(x) - f^*(\Delta)),$$
$$\forall\, f^* \in \mathcal{D}(\mathbb{A}^0), \quad \forall\, (x,y) \in \mathbb{R}^q \times F, \tag{A.2.32a}$$

$$\mathbb{A}^0 f^*(\Delta, y) := 0, \quad \forall\, f^* \in \mathcal{D}(\mathbb{A}^0), \quad y \in F. \tag{A.2.32b}$$

This defines a linear mapping $\mathbb{A}^0 : \mathcal{D}(\mathbb{A}^0) \subset \bar{C}(\mathbb{R}^{q*}) \to \bar{C}(\mathbb{R}^{q*} \times F)$. Next, fix an arbitrary $\bar{y} \in F$, and define a transition function η from $\mathbb{R}^{q*}$ to F by

$$\eta(x, \Gamma) := \delta_{\left(\frac{a(x)}{1+|x|^2}, \frac{b(x)}{1+|x|}, \lambda(x)\right)}(\Gamma), \quad x \in \mathbb{R}^q,\ \Gamma \in \mathcal{B}(F), \tag{A.2.33a}$$

$$\eta(\Delta, \Gamma) := \delta_{\bar{y}}(\Gamma), \quad \Gamma \in \mathcal{B}(F). \tag{A.2.33b}$$

Putting together (A.2.28), (A.2.31), (A.2.32) and (A.2.33), we get $\mathcal{G}^* \equiv \mathbb{A}_\eta^0$, where

$$\mathbb{A}_\eta^0 f^*(x) := \int_F \mathbb{A}^0 f^*(x,y)\eta(x, dy), \quad \forall\, x \in \mathbb{R}^{q*}, \quad \forall\, f^* \in \mathcal{D}(\mathbb{A}_\eta^0) := \mathcal{D}(\mathbb{A}^0). \tag{A.2.34}$$

We next check the conditions of Theorem A.2.107 for $\mathbb{A}^0$ and η given by (A.2.32) and (A.2.33): (i) From (A.2.27a) it follows that $\mathcal{D}(\mathbb{A}^0) := \mathcal{D}(\mathcal{C}^*)$ is closed under multiplication and separates points. (ii) Fix $y \in F$ and $\nu^* \in \mathcal{P}(\mathbb{R}^{q*})$, and define the linear operator $\mathcal{L}_y^*$ on $C(\mathbb{R}^{q*})$ by $\mathcal{D}(\mathcal{L}_y^*) := \mathcal{D}(\mathcal{C}^*)$ and

$$\mathcal{L}_y^* f^*(x) := \mathcal{L}_y(f^*|_{\mathbb{R}^q} - f^*(\Delta))(x), \quad \forall\, x \in \mathbb{R}^q, \qquad \mathcal{L}_y^* f^*(\Delta) := 0, \quad \forall\, f^* \in \mathcal{D}(\mathcal{L}_y^*).$$

The operator $\mathcal{L}_y$ given by (A.2.31) satisfies the positive maximum principle, thus Theorem 4.5.4 of Ethier and Kurtz [8] gives existence of a solution of the $D_{\mathbb{R}^{q*}}[0,\infty)$-martingale problem for $(\mathcal{L}_y^*, \nu^*)$. Then, for the linear operator $\mathbb{A}_y^0$ on $C(\mathbb{R}^{q*})$ given by $\mathcal{D}(\mathbb{A}_y^0) := \mathcal{D}(\mathcal{C}^*)$ and

$$\mathbb{A}_y^0 f^*(x) := \mathcal{L}_y^* f^*(x) - y_3(f^*(x) - f^*(\Delta)), \quad \forall\, x \in \mathbb{R}^{q*}, \quad \forall\, f^* \in \mathcal{D}(\mathbb{A}_y^0),$$

it follows from Theorem 4.10.2 of Ethier and Kurtz [8] that there exists a solution of the $D_{\mathbb{R}^{q*}}[0,\infty)$-martingale problem for $(\mathbb{A}_y^0, \delta_x)$, $\forall\, x \in \mathbb{R}^{q*}$, and therefore $\mathbb{A}_y^0$ is a pre-generator (see remark at foot of page 4 in [21]). (iii) From Remark 2.5 of Kurtz [21] there is a sequence $\{g_k\} \subset C_c^\infty(\mathbb{R}^q)$ such that the graph of $\mathcal{C}$ is included within the bp-closure of the linear span of $\{(g_k, \mathcal{C}g_k)\}$. Now define $g_{k,q}^*(x) := g_k(x) + q$, $\forall\, x \in \mathbb{R}^q$, $g_{k,q}^*(\Delta) := q$, $k = 1, 2, \ldots$, q rational. From (A.2.28) it follows that the countable set $\{g_{k,q}^*\}$ is a subset of $\mathcal{D}(\mathcal{G}^*)$ and the graph of $\mathcal{G}^*$ is contained in the bp-closure of the linear span of $\{(g_{k,q}^*, \mathcal{G}^* g_{k,q}^*)\}$. Since $\mathcal{G}^* = \mathbb{A}_\eta^0$, this checks condition (iii) of Theorem A.2.107, from which we conclude that uniqueness holds for the forward equation of $\mathcal{G}^*$, provided that uniqueness holds for the martingale problem for $\mathcal{G}^*$. To see that this is the case, observe from Theorem 8.1.7 of Ethier and Kurtz [8] that the martingale problem for $\mathcal{C}$ is well-posed, from which it follows that the $D_{\mathbb{R}^{q*}}[0,\infty)$-martingale problem for $\mathcal{C}^*$ is well-posed (see Lemma A.2.108), and hence Theorem 4.10.3 of [8] shows that the $D_{\mathbb{R}^{q*}}[0,\infty)$ martingale problem for $\mathcal{G}^*$ is well-posed. Now it follows from Theorem D.2.140 that uniqueness holds for the martingale problem for $\mathcal{G}^*$. □

Lemma A.2.108. *Under the conditions of Theorem 3.4.44 the $D_{\mathbb{R}^{q*}}[0,\infty)$ martingale problem for $\mathcal{C}^*$ defined in* (A.2.27) *is well posed.*

Proof. Let us first establish well-posedness for the martingale problem for $(\mathcal{C}^*, \delta_x)$ for $x \in \mathbb{R}^{q*}$. According to Theorem 8.1.7 of Ethier and Kurtz [8], for every $x \in \mathbb{R}^{q*}$ there exists a corlol solution $\{X_t\}$ of the martingale problem for $(\mathcal{C}, \delta_x)$. By inspection we see that $\{X_t\}$ also solves the martingale problem for $(\mathcal{C}^*, \delta_x)$. To show that this solution is unique, note that the conservativeness of $\mathcal{C}$ ensures that $(I_{\mathbb{R}^q}, 0)$ belongs to the bp-closure of $\mathcal{C}^*$ (see Definition D.2.139). Therefore, since by Theorem 8.1.7 of Ethier and Kurtz [8] the martingale problem for $(\mathcal{C}, \delta_x)$ is well-posed, the uniqueness for the $D_{\mathbb{R}^{q*}}[0, \infty)$ martingale problem for $(\mathcal{C}^*, \delta_x)$ follows by Theorem D.2.140.

The existence for the martingale problem for $(\mathcal{C}^*, \delta_\Delta)$ follows immediately, as the process $X_t \equiv \Delta$ is a solution. For uniqueness, fix some solution $\{X_t\}$, and observe that the conservativeness of $\mathcal{C}$ together with the dominated convergence theorem implies that the process $\{I_{\mathbb{R}^q}(X_t),\ t \in [0, \infty)\}$ is an $\{\mathcal{F}_t^X\}$-martingale. This gives

$$\mathrm{E}[(I_{\mathbb{R}^{2d}}(X_t) - I_{\mathbb{R}^{2d}}(X_0))^2] = \mathrm{E}[I_{\mathbb{R}^{2d}}(X_t)] - 2\,\mathrm{E}[I_{\mathbb{R}^{2d}}(X_t) I_{\mathbb{R}^{2d}}(X_0)] + \mathrm{E}[I_{\mathbb{R}^{2d}}(X_0)] = 0,$$

for all $t \in [0, \infty)$, so that

$$I_{\{\Delta\}}(X_t) = I_{\{\Delta\}}(X_0) = 1,\ a.s.,\ \forall t \in [0, \infty),$$

which in turn implies $P(\{X_t = \Delta,\ t \in [0, \infty)\}) = 1$.

To complete the proof we have to establish that the well-posedness of the martingale problem for $(\mathcal{C}^*, \delta_x)$ for $x \in \mathbb{R}^{q*}$ implies the well posedness of the martingale problem for $\mathcal{C}^*$. To that end let $P_x \in \mathcal{P}(C_{\mathbb{R}^{q*}}[0, \infty))$ be the solution of the martingale problem for $(\mathcal{C}^*, \delta_x)$, $x \in \mathbb{R}^{q*}$, and note that there exists a countable set $\{f_n\} \subset \mathcal{D}(\mathcal{C}^*)$ such that $\mathcal{C}^*$ is contained in the bp-closure of $\mathrm{span}\{(f_n, \mathcal{C}^* f_n)\}$. Thus, from Theorem D.2.138 it follows that the mapping $x \mapsto P_x$ is Borel measurable and that the martingale problem for $\mathcal{C}^*$ is well-posed. Therefore, $\mathcal{C}^*$ satisfies the conditions of Theorem 4.10.3 of Ethier and Kurtz [8], so that the result follows by a version of that theorem. □

Appendix B

Appendices for Chapter 4

Proof of Lemma 4.1.56. Define for every $\Psi \in C(S)$

$$U_n\Psi := \int_0^n [T_t\Psi - \bar{P}\Psi]\,dt, \quad n \in \mathbb{N}, \tag{B.0.1}$$

$$U\Psi := \int_0^\infty [T_t\Psi - \bar{P}\Psi]\,dt. \tag{B.0.2}$$

Notice that, by Lemma 1.1.4(a) of Ethier and Kurtz [8], Condition 4.1.54 ensures that the integral in (B.0.2) exists, and

$$U : C(S) \longrightarrow C(S). \tag{B.0.3}$$

Also, by definition, we have

$$\lim_{n\to\infty} \|U_n\Psi - U\Psi\| = 0, \quad \forall\Psi \in C(S). \tag{B.0.4}$$

Now $\|T_t\Psi\| \leq \|\Psi\|$ and $|\bar{P}\Psi| \leq \|\Psi\|$, so from (B.0.1) we have

$$\|U_n\Psi\| \leq \int_0^n \|T_t\Psi - \bar{P}\Psi\|\,dt \leq \int_0^n \|T_t\Psi\| + |\bar{P}\Psi|\,dt \leq 2n\|\Psi\|, \quad \forall\Psi \in C(S), \tag{B.0.5}$$

hence U_n is a bounded linear mapping on $C(S)$ for every $n \in \mathbb{N}$. Therefore, from (B.0.4) by the Uniform boundedness principle, U is a bounded linear mapping on

$C(S)$, and so for every $z \in S$ we have

$$|U\Psi(z)| \leq \|U\Psi\| \leq \|U\|\|\Psi\|, \quad \forall \Psi \in C(S),$$

which in turn implies that the mapping

$$\Psi \in C(S) \longrightarrow U\Psi(z) \in \mathbb{R}$$

is a bounded linear mapping for every $z \in S$. Thus, by the Riesz representation theorem (see Theorem 7.17 in Folland [9]), for every $z \in S$ there exists a finite signed Borel measure $\chi(z, \cdot)$ on S, such that

$$\|\chi(z,\cdot)\|_{TV} = \sup_{\|\Psi\|=1} |U\Psi(z)|, \quad \forall z \in S \tag{B.0.6}$$

and

$$U\Psi(z) = \int_S \Psi(z')\chi(z, dz'), \quad \forall z \in S, \forall \Psi \in C(S). \tag{B.0.7}$$

By (B.0.6)

$$\|\chi(z,\cdot)\|_{TV} \leq \|U\| < \infty, \quad \forall z \in S,$$

which together with (B.0.2), (B.0.3), and (B.0.7) establishes the first assertion.

To get the second part, fix some $\Psi \in C(S)$ such that $\bar{P}\Psi = 0$. Put

$$\Phi := \int_0^\infty T_t \Psi \, dt,$$

and note from the first part of the lemma that $\Phi \in C(S)$, and

$$\Phi(z) = \int_S \Psi(z')\chi(z, dz'), \quad \forall z \in S. \tag{B.0.8}$$

Let R_λ, $\lambda \in (0, \infty)$ be the resolvent of $\{T_t\}$:

$$R_\lambda \Psi = \int_0^\infty \exp(-\lambda t)\, T_t \Psi \, dt, \quad \lambda \in (0, \infty). \tag{B.0.9}$$

From (B.0.9) and the definition of Φ

$$\|R_\lambda \Psi - \Phi\| \le \int_0^\infty [1 - \exp(-\lambda t)] \|T_t \Psi\| \, dt,$$

so that Condition 4.1.54 and the dominated convergence theorem yield

$$\lim_{\lambda \to 0+} \|R_\lambda \Psi - \Phi\| = 0,$$

and thus

$$\lim_{\lambda \to 0+} (R_\lambda \Psi, \lambda R_\lambda \Psi - \Psi) = (\Phi, -\Psi). \tag{B.0.10}$$

Now $R_\lambda \Psi \in \mathcal{D}(\Omega)$, $\Omega R_\lambda \Psi = \lambda R_\lambda \Psi - \Psi$ (see page 11 of Ethier and Kurtz [8]), giving

$$(R_\lambda \Psi, \lambda R_\lambda \Psi - \Psi) \in \Omega, \quad \forall \lambda \in (0, \infty). \tag{B.0.11}$$

Since Ω is closed in $C(S) \times C(S)$ (see Corollary 1.1.6 of Ethier and Kurtz [8]), (B.0.10) and (B.0.11) yield $(\Phi, -\Psi) \in \Omega$, i.e.

$$\Phi \in \mathcal{D}(\Omega), \quad \Omega \Phi = -\Psi.$$

To get the third part, observe by Lemma D.3.147 that for every fixed $z \in S$ the mapping $f(\cdot, z)$ belongs to $C^1(\mathbb{R}^d)$, and

$$\partial_j f(x, z) = \int_S \partial_j g(x, z') \, \chi(z, dz'), \quad 1 \le j \le d, \ (x, z) \in \mathbb{R}^d \times S. \tag{B.0.12}$$

Thus, to complete the proof it suffices to show that for every $h \in C(\mathbb{R}^d \times S)$ the function

$$v(x, z) := \int_S h(x, z') \chi(z, dz'), \quad (x, z) \in \mathbb{R}^d \times S$$

belongs to $C(\mathbb{R}^d \times S)$, which, in particular, implies that $\partial_j f$ in (B.0.12) is a continuous function. To this end, let $C := \sup_z \|\chi(z, \cdot)\|_{TV}$ and let r be the metric on S. Fix some arbitrary $(x_0, z_0) \in \mathbb{R}^d \times S$, $\epsilon > 0$. We shall find $\delta > 0$ such that

$$\rho((x_0, z_0), (x, z)) \le \delta \Rightarrow |v(x_0, z_0) - v(x, z)| < \epsilon, \tag{B.0.13}$$

where

$$\rho((x_0, z_0), (x, z)) := r(z_0, z) + \max_{1 \le i \le d} |x_0^i - x^i|.$$

To this end, let $K_1 := \{x \in \mathbb{R}^d : \max_{1 \le i \le d} |x_0^i - x^i| \le 1\}$, and define

$$\hat{h}(x, z) := h(x, z), \quad \forall (x, z) \in K_1 \times S.$$

Observe that, since $K_1 \times S$ is compact, by the Stone-Weierstrass theorem there exist a pair of sequences $(f_i)_{i=1}^n \subset C(K_1)$, $(g_i)_{i=1}^n \subset C(S)$ such that

$$\left| \hat{h}(x, z) - \sum_{i=1}^n f_i(x) g_i(z) \right| < \frac{\epsilon}{3C}, \quad \forall (x, z) \in K_1 \times S \tag{B.0.14}$$

(see Exercise 4.68 of Folland [9]). Thus, for every $(x_1, z_1) \in K_1 \times S$ we have

$$\begin{aligned}
&\left| \int_S h(x_0, z')\chi(z_0, dz') - \int_S h(x_1, z')\chi(z_1, dz') \right| \\
\le &\left| \int_S [\hat{h}(x_0, z') - \sum_{i=1}^n f_i(x_0) g_i(z')] \chi(z_0, dz') \right| + \left| \int_S [\hat{h}(x_1, z') - \sum_{i=1}^n f_i(x_1) g_i(z')] \chi(z_1, dz') \right| \\
&+ \left| \sum_{i=1}^n \left[f_i(x_0) \int_S g_i(z')\chi(z_0, dz') - f(x_1) \int_S g_i(z')\chi(z_1, dz') \right] \right|.
\end{aligned} \tag{B.0.15}$$

Note by (B.0.14) that

$$\left| \int_S \left[\hat{h}(x_j, z') - \sum_{i=1}^n f_i(x_j) g_i(z') \right] \chi(z_j, dz') \right| \le \frac{\epsilon}{3C} C = \frac{\epsilon}{3}, \quad j = 0, 1. \tag{B.0.16}$$

Next, define

$$\Phi_i(z) := \int_S g_i(z')\chi(z, dz'), \quad \forall z \in S, \ 1 \le i \le n.$$

and observe by the first part of the lemma that $(\Phi_i)_{i=1}^n \subset C(S)$. Thus, we can find $\delta_1 > 0$ such that

$$\rho((x_0, z_0), (x, z)) \le \delta_1 \Rightarrow \left| \sum_{i=1}^n f_i(x_0)\Phi_i(z_0) - \sum_{i=1}^n f_i(x)\Phi_i(z) \right| < \frac{\epsilon}{3}, \quad \forall (x, z) \in K_1 \times S. \tag{B.0.17}$$

Since

$$\rho((x_0, z_0), (x, z)) \leq 1 \Rightarrow (x, z) \in K_1 \times S,$$

from (B.0.15), (B.0.16), and (B.0.17), with $\delta := \delta_1 \wedge 1$, we have

$$\rho((x_0, z_0), (x, z)) \leq \delta \Rightarrow \left| \int_S h(x_0, z')\chi(z_0, dz') - \int_S h(x, z')\chi(z, dz') \right| < \epsilon,$$

which is (B.0.13). □

To reduce notation in the proofs of Proposition 4.2.67 and Proposition 5.2.81, let

$$\mathcal{G}_t := \sigma\{Z_u, W_v,\ u \in [0, \infty),\ v \in [0, t]\} \vee \mathcal{N}(P), \quad t \in [0, \infty), \tag{B.0.18}$$

$$\mathcal{H}_t^\epsilon := \sigma\{Z_u^\epsilon, W_u,\ u \in [0, t]\} \vee \mathcal{N}(P), \quad t \in [0, \infty),\ \epsilon \in (0, 1]. \tag{B.0.19}$$

Proof of Proposition 4.2.67. Note that if we take the domain of $\mathcal{C}^\epsilon$ to be the collection of functions in the "product form", $f \otimes g$, $f \in C_c^2(\mathbb{R}^d)$, $g \in C(S)$, then the result follows easily by Itô's product rule. This suggests that one may try to use some kind of "density argument" in extending the domain beyond the functions in the product form, which is needed to accommodate the functions in (4.2.22). To make such a density argument work, however, one has to postulate boundedness of the generator $\mathcal{Q}$, a rather stringent condition. For that reason we adopt a different route, and establish the result from first principles.

Fix some $0 \leq t_1 < t_2 < \infty$. Using the independence of $\{\mathcal{F}_t^Z\}$ and $\{\mathcal{F}_t^W\}$, from (B.0.18) we see that $\{W_t\}$ is a $\{\mathcal{G}_t\}$-martingale and

$$\mathrm{E}[W_{t_2}^i W_{t_2}^j - W_{t_1}^i W_{t_1}^j | \mathcal{G}_{t_1}] = \mathrm{E}[W_{t_2}^i W_{t_2}^j - W_{t_1}^i W_{t_1}^j | \mathcal{F}_{t_1}^W] = \delta_{ij}(t_2 - t_1),$$

whence, by Lévy's theorem,

$$\{W_t\} \quad \text{is a } \{\mathcal{G}_t\}\text{-Wiener process.} \tag{B.0.20}$$

Also, since $\{\mathcal{H}_t^\epsilon\}$ is a subfiltration of $\{\mathcal{G}_t\}$, it follows that $\{X_t^\epsilon\}$ is adapted to $\{\mathcal{G}_t\}$. Thus, fixing arbitrary $0 \le u < v < \infty$, $z \in S$, by Itô's formula (recall $f \in \tilde{\mathcal{D}} \subset C_c^{2,0}(\mathbb{R}^d \times S)$), we get

$$\begin{aligned} f(X_v^\epsilon, z) - f(X_u^\epsilon, z) = \int_u^v \Bigg[\sum_i \left[\frac{1}{\epsilon} F^i(X_s^\epsilon, Z_s^\epsilon) + G^i(X_s^\epsilon, Z_s^\epsilon) \right] \partial_i f(X_s^\epsilon, z) \\ + \frac{1}{2} \sum_{i,j} [BB^T(X_s^\epsilon, Z_s^\epsilon)]^{ij} \partial_i \partial_j f(X_s^\epsilon, z) \Bigg] ds \\ + \int_u^v \sum_{i,j} B^{ij}(X_s^\epsilon, Z_s^\epsilon) \partial_i f(X_s^\epsilon, z)\, dW_s^j. \end{aligned} \tag{B.0.21}$$

Now $\partial_i f \in C_c(\mathbb{R}^d \times S)$, and thus by (B.0.20)

$$\left\{ \int_0^\cdot B^{ij}(X_s^\epsilon, Z_s^\epsilon) \partial_i f(X_s^\epsilon, z)\, dW_s^j \right\} \qquad \text{is a } \{\mathcal{G}_t\}\text{-martingale,}$$

so that conditioning in (B.0.21) gives

$$\begin{aligned} \mathrm{E}\Bigg[f(X_v^\epsilon, z) - f(X_u, z) - \int_u^v \Bigg(\sum_i [1/\epsilon\, F^i(X_s^\epsilon, Z_s^\epsilon) + G^i(X_s^\epsilon, Z_s^\epsilon)] \partial_i f(X_s^\epsilon, z) \\ + 1/2 \sum_{i,j} [BB^T(X_s^\epsilon, Z_s^\epsilon)]^{ij} \partial_i \partial_j f(X_s^\epsilon, z) \Bigg) ds \,\Bigg|\, \mathcal{G}_u \Bigg] = 0. \end{aligned} \tag{B.0.22}$$

Note that the argument of the conditional expectation in (B.0.22) is uniformly bounded and $\mathcal{F} \times \mathcal{B}(S)$-measurable, and that Z_v^ϵ is $\mathcal{G}_u$-measurable. Thus Lemma D.3.148 shows that

$$\begin{aligned} \mathrm{E}\Bigg[f(X_v^\epsilon, Z_v^\epsilon) - f(X_u^\epsilon, Z_v^\epsilon) - \int_u^v \Bigg(\sum_i [1/\epsilon\, F^i(X_s^\epsilon, Z_s^\epsilon) + G^i(X_s^\epsilon, Z_s^\epsilon)] \partial_i f(X_s^\epsilon, Z_v^\epsilon) \\ + 1/2 \sum_{i,j} [BB^T(X_s^\epsilon, Z_s^\epsilon)]^{ij} \partial_i \partial_j f(X_s^\epsilon, Z_v^\epsilon) \Bigg) ds \,\Bigg|\, \mathcal{G}_u \Bigg] = 0, \end{aligned}$$

hence, from (B.0.19), using the composition rule for conditioning, we get

$$\mathrm{E}\left[f(X_v^\epsilon, Z_v^\epsilon) - f(X_u^\epsilon, Z_v^\epsilon) - \int_u^v \left(\sum_i [1/\epsilon F^i(X_s^\epsilon, Z_s^\epsilon) + G^i(X_s^\epsilon, Z_s^\epsilon)]\partial_i f(X_s^\epsilon, Z_v^\epsilon) \right.\right.$$
$$\left.\left. +1/2\sum_{i,j}[BB^T(X_s^\epsilon, Z_s^\epsilon)]^{ij}\partial_i\partial_j f(X_s^\epsilon, Z_v^\epsilon)\right) ds \,\middle|\, \mathcal{H}_u^\epsilon\right] = 0. \quad \text{(B.0.23)}$$

To simplify notation, write

$$V(x, z_1, z_2) := \sum_i \left[\frac{1}{\epsilon}F^i(x, z_1) + G^i(x, z_1)\right] \partial_i f(x, z_2)$$
$$+ \frac{1}{2}\sum_{i,j}[BB^T(x, z_1)]^{ij}\partial_i\partial_j f(x, z_2), \quad \text{(B.0.24)}$$

so that (B.0.23) can be summarized as

$$\mathrm{E}\left[f(X_v^\epsilon, Z_v^\epsilon) - f(X_u^\epsilon, Z_v^\epsilon) - \int_u^v V(X_s^\epsilon, Z_s^\epsilon, Z_v^\epsilon)\, ds \,\middle|\, \mathcal{H}_u^\epsilon\right] = 0, \quad 0 \le u < v < \infty. \quad \text{(B.0.25)}$$

Next, if $\{T_t^\epsilon\}$ is the Feller semigroup on $C(S)$ corresponding to Markov process $\{Z_t^\epsilon\}$ defined by (4.1.3), then $\epsilon^{-2}\Omega$ is the generator of $\{T_t^\epsilon\}$ with domain $\mathcal{D}(\Omega)$. Since $f \in \tilde{\mathcal{D}}$, one has $f(x, \cdot) \in \mathcal{D}(\Omega)$, $\forall x \in \mathbb{R}^d$, and thus Proposition 4.1.7 of Ethier and Kurtz [8] shows that for every $x \in \mathbb{R}^d$

$$U_t^\epsilon(x) := f(x, Z_t^\epsilon) - \epsilon^{-2}\int_0^t \Omega[f(x, \cdot)](Z_s^\epsilon)\, ds, \quad t \in [0, \infty) \quad \text{is an } \mathcal{F}_t^{Z^\epsilon}\text{-martingale.} \quad \text{(B.0.26)}$$

Using the independence of $\mathcal{F}^W$ and $\mathcal{F}^Z$, we get

$$\mathrm{E}[U_v^\epsilon(x) - U_u^\epsilon(x)|\mathcal{H}_u^\epsilon] = \mathrm{E}[U_v^\epsilon(x) - U_u^\epsilon(x)|\mathcal{F}_u^{Z^\epsilon} \vee \mathcal{F}_u^W] = \mathrm{E}[U_v^\epsilon(x) - U_u^\epsilon(x)|\mathcal{F}_u^{Z^\epsilon}] = 0, \quad \text{(B.0.27)}$$

for every $x \in \mathbb{R}^d$. By the definition of $\tilde{\mathcal{D}}$ it follows that for every $t \in [0, \infty)$ the mapping

$$(x, \omega) \in \mathbb{R}^d \times \Omega \longmapsto U_t^\epsilon(x, \omega) \in \mathbb{R}$$

is uniformly bounded and $\mathcal{F}\times\mathcal{B}(\mathbb{R}^d)$-measurable, and thus, since X_u^ϵ is $\mathcal{H}_u^\epsilon$-measurable, from (B.0.27) and Lemma D.3.148, one has

$$\mathrm{E}[U_v^\epsilon(X_u^\epsilon) - U_u^\epsilon(X_u^\epsilon)|\mathcal{H}_u^\epsilon] = 0,$$

so by (B.0.26)

$$\mathrm{E}\left[f(X_u^\epsilon, Z_v^\epsilon) - f(X_u^\epsilon, Z_u^\epsilon) - \epsilon^{-2}\int_u^v \mathcal{Q}[f(X_u^\epsilon,\cdot)](Z_s^\epsilon)\,ds \,\middle|\, \mathcal{H}_u^\epsilon\right] = 0. \tag{B.0.28}$$

Next, fix some $0 \le s < t < \infty$ and let

$$t_k^n := s + \frac{k}{n}(t-s), \quad 0 \le k \le n,\ n \in \mathbb{N}.$$

Using $\mathcal{H}_s^\epsilon \subset \mathcal{H}_{t_k^n}^\epsilon$ we get

$$\begin{aligned}
\mathrm{E}[f(X_t^\epsilon, Z_t^\epsilon) - f(X_s^\epsilon, Z_s^\epsilon)\,|\,\mathcal{H}_s^\epsilon] &= \mathrm{E}\left[\sum_{k=0}^{n-1}[f(X_{t_{k+1}^n}^\epsilon, Z_{t_{k+1}^n}^\epsilon) - f(X_{t_k^n}^\epsilon, Z_{t_k^n}^\epsilon)] \,\middle|\, \mathcal{H}_s^\epsilon\right]\\
&= \mathrm{E}\left[\mathrm{E}\left[\sum_{k=0}^{n-1}[f(X_{t_{k+1}^n}^\epsilon, Z_{t_{k+1}^n}^\epsilon) - f(X_{t_k^n}^\epsilon, Z_{t_{k+1}^n}^\epsilon)] \,\middle|\, \mathcal{H}_{t_k^n}^\epsilon\right] \,\middle|\, \mathcal{H}_s^\epsilon\right]\\
&\quad + \mathrm{E}\left[\mathrm{E}\left[\sum_{k=0}^{n-1}[f(X_{t_k^n}^\epsilon, Z_{t_{k+1}^n}^\epsilon) - f(X_{t_k^n}^\epsilon, Z_{t_k^n}^\epsilon)] \,\middle|\, \mathcal{H}_{t_k^n}^\epsilon\right] \,\middle|\, \mathcal{H}_s^\epsilon\right].
\end{aligned} \tag{B.0.29}$$

From (B.0.25), (B.0.28), and (B.0.29) we now have

$$\begin{aligned}
\mathrm{E}[f(X_t^\epsilon, Z_t^\epsilon) - f(X_s^\epsilon, Z_s^\epsilon)\,|\,\mathcal{H}_s^\epsilon] &= \mathrm{E}\left[\sum_{k=0}^{n-1}\mathrm{E}\left[\int_{t_k^n}^{t_{k+1}^n} V(X_u^\epsilon, Z_u^\epsilon, Z_{t_{k+1}^n}^\epsilon)\,du \,\middle|\, \mathcal{H}_{t_k^n}^\epsilon\right] \,\middle|\, \mathcal{H}_s^\epsilon\right]\\
&\quad + \mathrm{E}\left[\sum_{k=0}^{n-1}\mathrm{E}\left[\int_{t_k^n}^{t_{k+1}^n} \epsilon^{-2}\mathcal{Q}[f(X_{t_k^n}^\epsilon,\cdot)](Z_u^\epsilon)\,du \,\middle|\, \mathcal{H}_{t_k^n}^\epsilon\right] \,\middle|\, \mathcal{H}_s^\epsilon\right]\\
&= \mathrm{E}\left[\sum_{k=0}^{n-1}\int_{t_k^n}^{t_{k+1}^n} V(X_u^\epsilon, Z_u^\epsilon, Z_{t_{k+1}^n}^\epsilon)\,du \,\middle|\, \mathcal{H}_s^\epsilon\right]\\
&\quad + \mathrm{E}\left[\sum_{k=0}^{n-1}\int_{t_k^n}^{t_{k+1}^n} \epsilon^{-2}\mathcal{Q}[f(X_{t_k^n}^\epsilon,\cdot)](Z_u^\epsilon)\,du \,\middle|\, \mathcal{H}_s^\epsilon\right].
\end{aligned} \tag{B.0.30}$$

Since $f \in \tilde{\mathcal{D}}$ and $\{X^\epsilon_t\}$ is a continuous process, we have

$$\operatorname*{bp-lim}_{n\to\infty} \sum_{k=0}^{n-1} \int_{t^n_k}^{t^n_{k+1}} \epsilon^{-2}\mathfrak{Q}[f(X^\epsilon_{t^n_k}, \cdot)](Z^\epsilon_u)\,du = \int_s^t \epsilon^{-2}\mathfrak{Q}[f(X^\epsilon_u, \cdot)](Z^\epsilon_u)\,du, \tag{B.0.31}$$

and, since $\{Z^\epsilon\}$ is right-continuous and $V \in C_c(\mathbb{R}^d \times S \times S)$, we have

$$\begin{aligned}\operatorname*{bp-lim}_{n\to\infty} \sum_{k=0}^{n-1} \int_{t^n_k}^{t^n_{k+1}} V(X^\epsilon_u, Z^\epsilon_u, Z^\epsilon_{t^n_{k+1}})\,du &= \int_s^t V(X^\epsilon_u, Z^\epsilon_u, Z^\epsilon_u)\,du \\ &= \int_s^t \Bigg[\sum_i [1/\epsilon\, F^i(X^\epsilon_u, Z^\epsilon_u) + G^i(X^\epsilon_u, Z^\epsilon_u)]\partial_i f(X^\epsilon_u, Z^\epsilon_u) \\ &\quad +1/2 \sum_{i,j} [BB^T(X^\epsilon_u, Z^\epsilon_u)]^{ij}\partial_i\partial_j f(X^\epsilon_u, Z^\epsilon_u)\Bigg]\,du. \end{aligned} \tag{B.0.32}$$

Using (B.0.30), (B.0.31), (B.0.32), and the dominated convergence theorem for conditional expectation, we get

$$\begin{aligned} \mathrm{E}[f(X^\epsilon_t, Z^\epsilon_t) - f(X^\epsilon_s, Z^\epsilon_s) \mid \mathcal{H}^\epsilon_s] &= \mathrm{E}\Bigg[\int_s^t \frac{1}{\epsilon^2}\mathfrak{Q}[f(X^\epsilon_u, \cdot)](Z^\epsilon_u) \\ &+ \frac{1}{\epsilon}\sum_i F^i(X^\epsilon_u, Z^\epsilon_u)\partial_i f(X^\epsilon_u, Z^\epsilon_u) + \sum_i G^i(X^\epsilon_u, Z^\epsilon_u)\partial_i f(X^\epsilon_u, Z^\epsilon_u) \\ &+ \frac{1}{2}\sum_{i,j}[BB^T(X^\epsilon_u, Z^\epsilon_u)]^{ij}\partial_i\partial_j f(X^\epsilon_u, Z^\epsilon_u)\Bigg]\,du \,\Bigg|\, \mathcal{H}^\epsilon_s\Bigg] = \mathrm{E}\left[\int_s^t \mathfrak{C}^\epsilon f(X^\epsilon_u, Z^\epsilon_u)\,du \,\middle|\, \mathcal{H}^\epsilon_s\right], \end{aligned}$$

which by (4.2.21) in turn implies $\mathrm{E}[M^{\epsilon,f}_t \mid \mathcal{H}^\epsilon_s] = M^{\epsilon,f}_s$, completing the proof. □

Proof of Proposition 4.2.68. Fix $\epsilon \in (0,1]$ and $\phi \in \mathcal{D}(\mathcal{L}) := C^\infty_c(\mathbb{R}^d)$ (recall (4.1.12)), and put

$$g^\phi_1(x,z) := \sum_{i=1}^d F^i(x,z)\partial_i\phi(x), \quad (x,z) \in \mathbb{R}^d \times S \tag{B.0.33}$$

$$f^\phi_1(x,z) := \int_S g^\phi_1(x,z')\chi(z,dz'), \quad (x,z) \in \mathbb{R}^d \times S. \tag{B.0.34}$$

Observe by Condition 4.1.51 and Lemma 4.1.56(iii) that

$$g^\phi_1, f^\phi_1 \in C^{3,0}_c(\mathbb{R}^d \times S). \tag{B.0.35}$$

Next, put

$$g_2^\phi(x,z) := \sum_{i=1}^d F^i(x,z)\partial_i f_1^\phi(x,z), \quad (x,z) \in \mathbb{R}^d \times S, \tag{B.0.36}$$

$$\bar{g}_2^\phi(x) := \int_S g_2^\phi(x,z')\,\bar{P}(dz'), \quad x \in \mathbb{R}^d, \tag{B.0.37}$$

$$f_2^\phi(x,z) := \int_S [g_2^\phi(x,z') - \bar{g}_2^\phi(x)]\chi(z,dz'), \quad (x,z) \in \mathbb{R}^d \times S. \tag{B.0.38}$$

From (B.0.35) we know that $\partial_i f_1^\phi \in C_c^{2,0}(\mathbb{R}^d \times S)$, hence Condition 4.1.51 and (B.0.35) show that $g_2^\phi \in C_c^{2,0}(\mathbb{R}^d \times S)$, and thus, by Lemma D.3.147, $\bar{g}_2^\phi \in C_c^2(\mathbb{R}^d)$. It therefore follows that $g_2^\phi - \bar{g}_2^\phi \in C_c^{2,0}(\mathbb{R}^d \times S)$, so by Lemma 4.1.56(3) we have $f_2^\phi \in C_c^{2,0}(\mathbb{R}^d \times S)$. To summarize:

$$g_2^\phi - \bar{g}_2^\phi,\ f_2^\phi \in C_c^{2,0}(\mathbb{R}^d \times S). \tag{B.0.39}$$

Let

$$g_3^\phi(x,z) := \sum_{i=1}^d G^i(x,z)\partial_i\phi(x) + \frac{1}{2}\sum_{i,j=1}^d [BB^T(x,z)]^{ij}\partial_i\partial_j\phi(x), \quad (x,z) \in \mathbb{R}^d \times S, \tag{B.0.40}$$

$$\bar{g}_3^\phi(x) := \int_S g_3^\phi(x,z')\,\bar{P}(dz'), \quad x \in \mathbb{R}^d, \tag{B.0.41}$$

$$f_3^\phi(x,z) := \int_S [g_3^\phi(x,z') - \bar{g}_3^\phi(x)]\chi(z,dz'), \quad (x,z) \in \mathbb{R}^d \times S. \tag{B.0.42}$$

From Condition 4.1.51 and (B.0.40) we know that $g_3^\phi \in C_c^{2,0}(\mathbb{R}^d \times S)$, so that Lemma D.3.147 gives $\bar{g}_3^\phi \in C_c^2(\mathbb{R}^d)$. Hence by Lemma 4.1.56(3) from (B.0.42) we have $f_3^\phi \in C_c^{2,0}(\mathbb{R}^d \times S)$, i.e.

$$g_3^\phi - \bar{g}_3^\phi,\ f_3^\phi \in C_c^{2,0}(\mathbb{R}^d \times S). \tag{B.0.43}$$

Now define the "perturbed test function" $f^{\epsilon,\phi}$ as in (4.2.22). We verify that $f^{\epsilon,\phi}$ belongs to $\tilde{\mathcal{D}}$, and that $\mathfrak{C}^\epsilon f^{\epsilon,\phi}$ has the form given in (4.2.23). Observe by (B.0.35), (B.0.39), (B.0.43), and (4.2.22) that we have

$$f^{\epsilon,\phi} \in C_c^{2,0}(\mathbb{R}^d \times S). \tag{B.0.44}$$

Also, from Condition 4.1.55 and (B.0.33) we know that

$$\int_S g_1^\phi(x,z')\,\bar{P}(dz') = 0, \quad \forall x \in \mathbb{R}^d,$$

hence (B.0.34) and Lemma 4.1.56(2) show that

$$f_1^\phi(x,\cdot) \in \mathcal{D}(\Omega), \quad \forall x \in \mathbb{R}^d,$$

with

$$\Omega[f_1^\phi(x,\cdot)](z) = -g_1^\phi(x,z), \quad \forall (x,z) \in \mathbb{R}^d \times S. \tag{B.0.45}$$

From (B.0.37) we see that

$$\int_S [g_2^\phi(x,z') - \bar{g}_2^\phi(x)]\,\bar{P}(dz') = 0, \quad \forall x \in \mathbb{R}^d,$$

so (B.0.38) and Lemma 4.1.56(2) imply

$$f_2^\phi(x,\cdot) \in \mathcal{D}(\Omega), \quad \forall x \in \mathbb{R}^d,$$

with

$$\Omega[f_2^\phi(x,\cdot)](z) = \bar{g}_2^\phi(x) - g_2^\phi(x,z), \quad \forall (x,z) \in \mathbb{R}^d \times S. \tag{B.0.46}$$

Similarly, from (B.0.41) we have

$$\int_S [g_3^\phi(x,z') - \bar{g}_3^\phi(x)]\,\bar{P}(dz') = 0, \quad \forall x \in \mathbb{R}^d,$$

thus (B.0.42) and Lemma 4.1.56(2) yield

$$f_3^\phi(x,\cdot) \in \mathcal{D}(\Omega), \quad \forall x \in \mathbb{R}^d,$$

and

$$\Omega[f_3^\phi(x,\cdot)](z) = \bar{g}_3^\phi(x) - g_3^\phi(x,z), \quad \forall (x,z) \in \mathbb{R}^d \times S. \tag{B.0.47}$$

We therefore see that $f_1^\phi(x,\cdot), f_2^\phi(x,\cdot), f_3^\phi(x,\cdot) \in \mathcal{D}(\Omega)$, $\forall x \in \mathbb{R}^d$, hence (4.2.22) and the fact that $1 \in \mathcal{D}(\Omega)$ with $\Omega 1 = 0$ (since $\{T_t\}$ is conservative by Condition 4.1.52) shows that

$$f^{\epsilon,\phi}(x,\cdot) \in \mathcal{D}(\Omega), \quad \forall x \in \mathbb{R}^d. \tag{B.0.48}$$

Now, (4.2.22), (B.0.45), (B.0.46), (B.0.47), (B.0.48), and $\mathfrak{Q}1 \equiv 0$ give

$$\begin{aligned}\mathfrak{Q}[f^{\epsilon,\phi}(x,\cdot)](z) &= \epsilon\mathfrak{Q}[f_1^{\epsilon,\phi}(x,\cdot)](z) + \epsilon^2\mathfrak{Q}[f_2^{\epsilon,\phi}(x,\cdot)](z) + \epsilon^2\mathfrak{Q}[f_3^{\epsilon,\phi}(x,\cdot)](z) \\ &= -\epsilon g_1^{\phi}(x,z) + \epsilon^2[\bar{g}_2^{\phi}(x) - g_2^{\phi}(x,z)] + \epsilon^2[\bar{g}_3^{\phi}(x) - g_3^{\phi}(x,z)]. \quad \text{(B.0.49)}\end{aligned}$$

By (B.0.35), (B.0.39), (B.0.43), and (B.0.49) we see that the mapping $(x,z) \mapsto \mathfrak{Q}[f^{\epsilon,\phi}(x,\cdot)](z)$ defines a member of $C_c^{2,0}(\mathbb{R}^d \times S)$, which, together with (4.2.19), implies

$$f^{\epsilon,\phi} \in \tilde{\mathcal{D}}. \quad \text{(B.0.50)}$$

We now check to see that $\mathfrak{C}^\epsilon f^{\epsilon,\phi}$ has the form (4.2.23): by (4.2.20), (4.2.22) and (B.0.49)

$$\begin{aligned}\mathfrak{C}^\epsilon f^{\epsilon,\phi}(x,z) = &-\frac{1}{\epsilon}g_1^{\phi}(x,z) + \bar{g}_2^{\phi}(x) - g_2^{\phi}(x,z) + \bar{g}_3^{\phi}(x) - g_3^{\phi}(x,z) \\ &+\frac{1}{\epsilon}\sum_i F^i(x,z)\partial_i\phi(x) + \sum_i F^i(x,z)\partial_i f_1^{\phi}(x,z) \\ &+\epsilon\sum_i F^i(x,z)\partial_i f_2^{\phi}(x,z) + \epsilon\sum_i F^i(x,z)\partial_i f_3^{\phi}(x,z) \\ &+\frac{1}{2}\sum_{i,j}[BB^T(x,z)]^{ij}\partial_i\partial_j\phi(x) + \frac{\epsilon}{2}\sum_{i,j}[BB^T(x,z)]^{ij}\partial_i\partial_j f_1^{\phi}(x,z) \\ &+\frac{\epsilon^2}{2}\sum_{i,j}[BB^T(x,z)]^{ij}\partial_i\partial_j f_2^{\phi}(x,z) + \frac{\epsilon^2}{2}\sum_{i,j}[BB^T(x,z)]^{ij}\partial_i\partial_j f_3^{\phi}(x,z) \\ &+\sum_i G^i(x,z)\partial_i\phi(x) + \epsilon\sum_i G^i(x,z)\partial_i f_1^{\phi}(x,z) \\ &+\epsilon^2\sum_i G^i(x,z)\partial_i f_2^{\phi}(x,z) + \epsilon^2\sum_i G^i(x,z)\partial_i f_3^{\phi}(x,z). \quad \text{(B.0.51)}\end{aligned}$$

Notice that in view of (B.0.35), (B.0.39), and (B.0.43) all functions obtained by taking partial x-derivatives in (B.0.51) belong to $C_c(\mathbb{R}^d \times S)$. It therefore follows that the function

$$\begin{aligned}\gamma^{\phi}(x,z,\epsilon) := &\sum_i F^i(x,z)\partial_i f_2^{\phi}(x,z) + \sum_i F^i(x,z)\partial_i f_3^{\phi}(x,z)\\ &+ \frac{1}{2}\sum_{i,j}[BB^T(x,z)]^{ij}\partial_i\partial_j f_1^{\phi}(x,z) + \frac{\epsilon}{2}\sum_{i,j}[BB^T(x,z)]^{ij}\partial_i\partial_j f_2^{\phi}(x,z)\\ &+ \frac{\epsilon}{2}\sum_{i,j}[BB^T(x,z)]^{ij}\partial_i\partial_j f_3^{\phi}(x,z) + \sum_i G^i(x,z)\partial_i f_1^{\phi}(x,z)\\ &+ \epsilon\sum_i G^i(x,z)\partial_i f_2^{\phi}(x,z) + \sum_i G^i(x,z)\partial_i f_3^{\phi}(x,z) \end{aligned}\tag{B.0.52a}$$

is jointly continuous on $\mathbb{R}^d \times S \times (0,1]$, and

$$\sup_{(x,z,\epsilon)\in\mathbb{R}^d\times S\times(0,1]} |\gamma^{\phi}(x,z,\epsilon)| < \infty. \tag{B.0.52b}$$

Thus, from (B.0.51) and (B.0.52a) we can write

$$\begin{aligned}\mathcal{C}^{\epsilon} f^{\epsilon,\phi}(x,z) =&1/\epsilon\left[\sum_i F^i(x,z)\partial_i\phi(x) - g_1^{\phi}(x,z)\right] + \left[\sum_i F^i(x,z)\partial_i f_1^{\phi}(x,z) - g_2^{\phi}(x,z)\right]\\ &+\left[\sum_i G^i(x,z)\partial_i\phi(x) + 1/2\sum_{i,j}[BB^T(x,z)]^{ij}\partial_i\partial_j\phi(x) - g_3^{\phi}(x,z)\right]\\ &+\bar{g}_2^{\phi}(x) + \bar{g}_3^{\phi}(x) + \epsilon\gamma^{\phi}(x,z,\epsilon), \quad \forall (x,z)\in\mathbb{R}^d\times S.\end{aligned}$$

Since the expressions in brackets on the right-hand side are identically zero (recall (B.0.33), (B.0.36), and (B.0.40)), this becomes

$$\mathcal{C}^{\epsilon} f^{\epsilon,\phi}(x,z) = \bar{g}_2^{\phi}(x) + \bar{g}_3^{\phi}(x) + \epsilon\gamma^{\phi}(x,z,\epsilon), \quad \forall (x,z,\epsilon)\in\mathbb{R}^d\times S\times(0,1]. \tag{B.0.53}$$

It remains to evaluate $\bar{g}_2^{\phi}(x)$ and $\bar{g}_3^{\phi}(x)$ in order to show that the right side of (B.0.53) is of the form (4.2.18). Considering $\bar{g}_2^{\phi}(x)$, one sees from (B.0.34), (B.0.35), and Lemma 4.1.56(3) that

$$\begin{aligned}\partial_i f_1^{\phi}(x,z) &= \int_S \partial_i g_1^{\phi}(x,z')\chi(z,dz') = \int_S \partial_i\left[\sum_j F^j(x,z')\partial_j\phi(x)\right]\chi(z,dz')\\ &= \sum_j\left[\int_S \partial_i F^j(x,z')\chi(z,dz')\right]\partial_j\phi(x) + \sum_j\left[\int_S F^j(x,z')\chi(z,dz')\right]\partial_i\partial_j\phi(x).\end{aligned}$$

Combining this with (B.0.36) gives

$$g_2^\phi(x,z) = \sum_{i,j} \left[\int_S \partial_i F^j(x,z')\chi(z,dz')\right] F^i(x,z)\partial_j\phi(x) + \sum_{i,j} \left[\int_S F^j(x,z')\chi(z,dz')\right] F^i(x,z)\partial_i\partial_j\phi(x),$$

which, by (B.0.37), gives

$$\bar{g}_2^\phi(x) = \sum_j \int_S \left[\sum_i \int_S \partial_i F^j(x,z')\chi(z,dz')F^i(x,z)\right] \bar{P}(dz)\partial_j\phi(x) + \sum_{i,j} \int_S \left[\int_S F^j(x,z')\chi(z,dz')\right] F^i(x,z)\,\bar{P}(dz)\,\partial_i\partial_j\phi(x). \tag{B.0.54}$$

From (B.0.40) and (B.0.41),

$$\bar{g}_3^\phi(x) = \sum_i \int_S G^i(x,z)\,\bar{P}(dz)\partial_i\phi(x) + 1/2\sum_{i,j}\int_S [BB^T(x,z)]^{ij}\,\bar{P}(dz)\partial_i\partial_j\phi(x) \tag{B.0.55}$$

Putting together (4.1.10), (4.1.11), (B.0.53), (B.0.54), and (B.0.55) gives

$$\mathfrak{C}^\epsilon f^{\epsilon,\phi}(x,z) = \mathfrak{L}\phi(x) + \epsilon\gamma^\phi(x,z,\epsilon), \quad \forall (x,z,\epsilon) \in \mathbb{R}^d \times S \times (0,1], \tag{B.0.56}$$

which is of the form (4.2.23). □

Appendix C

Appendices for Chapter 5

C.1 Appendix for Section 5.2

Fact C.1.109. *Let E be a metric space. For an arbitrary $f \in \bar{C}(E)$ the mapping*

$$\Psi : \nu_{\cdot} \in C_{\mathcal{P}(E)}[0,T] \longmapsto \int_0^{\cdot} \nu_u f \, du \in C_{\mathbb{R}}[0,T]$$

is continuous.

Proof. Introduce the mappings

$$\Psi_1 : \nu_{\cdot} \in C_{\mathcal{P}(E)}[0,T] \longmapsto \nu_{\cdot} f \in C_{\mathbb{R}}[0,T],$$
$$\Psi_2 : x(\cdot) \in C_{\mathbb{R}}[0,T] \longmapsto \int_0^{\cdot} x(u) \, du \in C_{\mathbb{R}}[0,T].$$

Since $f \in \bar{C}(E)$ and $[0,T]$ is compact, the first mapping is continuous by Theorem XII.2.2(2) of Dugundji [7]. On the other hand, since

$$\left| \int_0^t x_1(u) \, du - \int_0^t x_2(u) \, du \right| \le \int_0^t |x_1(u) - x_2(u)| \, du \le T \|x_1 - x_2\|,$$

for all $t \in [0,T]$, $x_1, x_2 \in C_{\mathbb{R}}[0,T]$, the second mapping is continuous as well. Therefore, Ψ is continuous as a composition of two continuous mappings. $\square$

Proof of Lemma 5.2.81. Fix some $\epsilon \in (0,1]$, $1 \leq i \leq r$, $f \in \tilde{\mathcal{D}}$. We shall show that

$$f(X^\epsilon_t, Z^\epsilon_t) - W^i_t \int_0^t \mathcal{C}^\epsilon f(X^\epsilon_s, Z^\epsilon_s)\, ds - \int_0^t \sum_j \partial_j f(X^\epsilon_s, Z^\epsilon_s) B^{ji}(X^\epsilon_s)\, ds, \quad t \in [0,T]$$

is an $\{\mathcal{H}^\epsilon_t\}$-martingale (see (B.0.19)). Since

$$\mathrm{E}\left[\sup_{u\in[s,t]} |W^i_u|\right] < \infty, \quad 0 \leq s < t < \infty, \tag{C.1.1}$$

it follows that

$$\mathrm{E}\left[\int_0^t |W^i_s\, \mathcal{C}^\epsilon f(X^\epsilon_s, Z^\epsilon_s)|\, ds\right] < \infty, \quad \forall t \in [0,T].$$

Therefore, by Lemma D.3.142 we see that it suffices to establish that

$$f(X^\epsilon_t, Z^\epsilon_t) W^i_t - \int_0^t W^i_u\, \mathcal{C}^\epsilon f(X^\epsilon_u, Z^\epsilon_u)\, du - \int_0^t \sum_j \partial_j f(X^\epsilon_u, Z^\epsilon_u) B^{ji}(X^\epsilon_u)\, du \quad t \in [0,T] \tag{C.1.2}$$

is an $\{\mathcal{H}^\epsilon_t\}$-martingale.

Fix some $0 \leq s < t < \infty$. To facilitate the exposition, we first complete the following two technical steps:

Step 1. Define the mapping $V : \Omega \times [0,T] \times \mathbb{R}^d \times S \times S \to \mathbb{R}$ by

$$\begin{aligned} V_u(\omega, x, z_1, z_2) := {} & W^i_u(\omega) \sum_j \left[1/\epsilon\, F^j(x,z_1) + G^j(x,z_1)\right] \partial_j f(x, z_2) \\ & + 1/2 \sum_{k,j} [BB^T(x)]^{kj} \partial_k \partial_j f(x, z_2) \\ & + \sum_k B^{ki}(x) \partial_k f(x, z_2), \quad (t, \omega, x, z_1, z_2) \in [0,T] \times \Omega \times \mathbb{R}^d \times S \times S, \end{aligned}$$

and show that (suppressing the ω-dependence of V)

$$\mathrm{E}\left[f(X^\epsilon_t, Z^\epsilon_t) W^i_t - f(X^\epsilon_s, Z^\epsilon_t) W^i_s - \int_s^t V_u(X^\epsilon_u, Z^\epsilon_u, Z^\epsilon_t)\, du \,\middle|\, \mathcal{H}^\epsilon_s \right] = 0. \tag{C.1.3}$$

To complete this step, note by the Itô's product rule that from (B.0.21) we get

$$f(X_t^\epsilon, z)W_t^i - f(X_s^\epsilon, z)W_s^i = \int_s^t V_u(X_u^\epsilon, Z_u^\epsilon, z)\,du \;+\; \{\mathcal{G}_t\}\text{-martingale}, \quad \forall z \in S$$

Conditioning together with (B.0.20) gives

$$\mathrm{E}\left[f(X_t^\epsilon, z)W_t^i - f(X_s^\epsilon, z)W_s^i - \int_s^t V_u(X_u^\epsilon, Z_u^\epsilon, z)\,du \,\middle|\, \mathcal{G}_s \right] = 0, \quad \forall z \in S. \quad \text{(C.1.4)}$$

Since $f \in C_c^{2,0}(\mathbb{R}^d \times S)$, it follows that there exists $C_1 \in [0, \infty)$ such that for every $z \in S$

$$\mathrm{E}\left[\left| f(X_t^\epsilon, z)W_t^i - f(X_s^\epsilon, z)W_s^i - \int_s^t V_u(X_u^\epsilon, Z_u^\epsilon, z)\,du \right| \right] \leq C_1 \,\mathrm{E}\left[\sup_{u \in [s,t]} |W_u^i| \right]. \quad \text{(C.1.5)}$$

Therefore, using (C.1.4), (C.1.5), and the fact that Z_t^ϵ is $\mathcal{G}_s$-measurable, by Lemma D.3.148, we have

$$\mathrm{E}\left[f(X_t^\epsilon, Z_t^\epsilon)W_t^i - f(X_s^\epsilon, Z_t^\epsilon)W_s^i - \int_s^t V_u(X_u^\epsilon, Z_u^\epsilon, Z_t^\epsilon)\,du \,\middle|\, \mathcal{G}_s \right] = 0,$$

whence (C.1.3) follows by conditioning on $\mathcal{H}_s^\epsilon$.

Step 2. Establish

$$\mathrm{E}\left[W_s^i f(X_s^\epsilon, Z_t^\epsilon) - W_s^i f(X_s^\epsilon, Z_s^\epsilon) - \frac{1}{\epsilon^2} W_s^i \int_s^t \mathcal{Q}[f(X_s^\epsilon, \cdot)](Z_u^\epsilon)\,du \,\middle|\, \mathcal{H}_s^\epsilon \right] = 0. \quad \text{(C.1.6)}$$

To this end, for every $n \in \mathbb{N}, x \in \mathbb{R}^d, y \in \mathbb{R}, t \in [0, T]$ put

$$\rho_t^n(x, y) := y I_{[-n,n]}(y) f(x, Z_t^\epsilon) - \frac{1}{\epsilon^2} \int_0^t y I_{[-n,n]}(y) \mathcal{Q}[f(x, \cdot)](Z_s^\epsilon)\,ds,$$

and notice that by Proposition 4.1.7 of Ethier and Kurtz [8] we have

$$\mathrm{E}[\rho_t^n(x, y) - \rho_s^n(x, y) | \mathcal{H}_s^\epsilon] = 0, \quad \forall x \in \mathbb{R}^d,\ y \in \mathbb{R},\ \forall n \in \mathbb{N}. \quad \text{(C.1.7)}$$

Thus, in view of boundedness of $\rho_t^n(x, y) - \rho_s^n(x, y)$ on $\mathbb{R}^d \times \mathbb{R} \times \Omega$ and $\mathcal{H}_s^\epsilon$-measurability of (X_s^ϵ, W_s^i), by Lemma D.3.148, we have

$$\mathrm{E}[\rho_t^n(X_s^\epsilon, W_s^i) - \rho_s^n(X_s^\epsilon, W_s^i) | \mathcal{H}_s^\epsilon] = 0,$$

whence (C.1.6) follows by taking the limit $n \to \infty$ and using the dominated convergence theorem for conditional expectations (recall (C.1.1)).

Fix now some arbitrary $0 \le u < v < \infty$ and let

$$t_k^n := u + \frac{k}{n}(v-u), \quad 0 \le k \le n,\ n \in \mathbb{N}.$$

Using the fact that $\mathcal{H}_u^\epsilon \subset \mathcal{H}_{t_k}^\epsilon$, we can write

$$\begin{aligned}
&\mathrm{E}[f(X_v^\epsilon, Z_v^\epsilon)W_v^i - f(X_u^\epsilon, Z_u^\epsilon)W_u^i|\mathcal{H}_u^\epsilon] \\
&\quad = \mathrm{E}\left[\sum_{k=0}^{n-1} \mathrm{E}[f(X_{t_{k+1}^n}^\epsilon, Z_{t_{k+1}^n}^\epsilon)W_{t_{k+1}^n}^i - f(X_{t_k^n}^\epsilon, Z_{t_k^n}^\epsilon)W_{t_k^n}^i \mid \mathcal{H}_{t_k^n}^\epsilon] \,\middle|\, \mathcal{H}_u^\epsilon\right] \\
&\quad = \mathrm{E}\left[\sum_{k=0}^{n-1} \mathrm{E}[f(X_{t_{k+1}^n}^\epsilon, Z_{t_{k+1}^n}^\epsilon)W_{t_{k+1}^n}^i - f(X_{t_k^n}^\epsilon, Z_{t_{k+1}^n}^\epsilon)W_{t_k^n}^i \mid \mathcal{H}_{t_k}^\epsilon] \,\middle|\, \mathcal{H}_u^\epsilon\right] \\
&\quad + \mathrm{E}\left[\sum_{k=0}^{n-1} \mathrm{E}[f(X_{t_k^n}^\epsilon, Z_{t_{k+1}^n}^\epsilon)W_{t_k^n}^i - f(X_{t_k^n}^\epsilon, Z_{t_k^n}^\epsilon)W_{t_k^n}^i \mid \mathcal{H}_{t_k}^\epsilon] \,\middle|\, \mathcal{H}_u^\epsilon\right],
\end{aligned}$$

which, together with (C.1.3), (C.1.6), and the composition rule for conditioning, gives

$$\begin{aligned}
&\mathrm{E}[f(X_v^\epsilon, Z_v^\epsilon)W_v^i - f(X_u^\epsilon, Z_u^\epsilon)W_u^i \mid \mathcal{H}_u^\epsilon] \\
&\quad = \sum_{k=0}^{n-1} \mathrm{E}\left[\int_{t_k^n}^{t_{k+1}^n} V_s(X_s^\epsilon, Z_s^\epsilon, Z_{t_{k+1}^n}^\epsilon) + \frac{1}{\epsilon^2} W_{t_k^n}^i \Omega[f(X_{t_k^n}^\epsilon, \cdot)](Z_s^\epsilon)\, ds \,\middle|\, \mathcal{H}_u^\epsilon\right].
\end{aligned} \tag{C.1.8}$$

By the continuity of $\{X_t^\epsilon\}$ and $\{W_t^i\}$ we have

$$\operatorname*{bp-lim}_{n\to\infty} \sum_{k=0}^{n-1} \int_{t_k^n}^{t_{k+1}^n} \frac{1}{\epsilon^2} W_{t_k^n}^i \Omega[f(X_{t_k^n}^\epsilon, \cdot)](Z_s^\epsilon)\, ds = \frac{1}{\epsilon^2}\int_u^v W_s^i \Omega[f(X_s^\epsilon, \cdot)](Z_s^\epsilon)\, ds. \tag{C.1.9}$$

Similarly, the right-continuity of $\{Z_t^\epsilon\}$ implies

$$\operatorname*{bp-lim}_{n\to\infty} \sum_{k=1}^{n-1} \int_{t_k^n}^{t_{k+1}^n} V_s(X_s^\epsilon, Z_s^\epsilon, Z_{t_{k+1}^n}^\epsilon)\, ds = \int_u^v V_s(X_s^\epsilon, Z_s^\epsilon, Z_s^\epsilon)\, ds. \tag{C.1.10}$$

From (C.1.1), (C.1.9), and the fact that $\Omega \subset C_c(\mathbb{R}^d \times S) \times C_c(\mathbb{R}^d \times S)$, by the dominated convergence theorem for conditional expectations, we get

$$\lim_{n\to\infty} \mathrm{E}\left[\sum_{k=0}^{n-1}\int_{t_k^n}^{t_{k+1}^n} \frac{1}{\epsilon^2} W^i_{t_k^n} \mathcal{Q}[f(X^\epsilon_{t_k^n},\cdot)](Z^\epsilon_s)\,ds \,\middle|\, \mathcal{H}^\epsilon_u\right]$$
$$= \mathrm{E}\left[\frac{1}{\epsilon^2}\int_u^v W^i_s \mathcal{Q}[f(X^\epsilon_s,\cdot)](Z^\epsilon_s)\,ds \,\middle|\, \mathcal{H}^\epsilon_u\right], \quad \text{(C.1.11)}$$

and, by the same argument, from (C.1.1), (C.1.10), we have

$$\lim_{n\to\infty}\sum_{k=1}^{n-1} \mathrm{E}\left[\int_{t_k^n}^{t_{k+1}^n} V_s(X^\epsilon_s, Z^\epsilon_s, Z^\epsilon_{t^n_{k+1}})\,ds \,\middle|\, \mathcal{H}^\epsilon_u\right] = \mathrm{E}\left[\int_u^v V_s(X^\epsilon_s, Z^\epsilon_s, Z^\epsilon_s)\,ds \,\middle|\, \mathcal{H}^\epsilon_u\right].$$
$$\text{(C.1.12)}$$

Taking the limit in (C.1.8), and using the definition of V, (C.1.11), and (C.1.12), we get

$$\mathrm{E}\left[f(X^\epsilon_v, Z^\epsilon_v)W^i_v - f(X^\epsilon_u, Z^\epsilon_u)W^i_u - \int_u^v W^i_s \mathcal{C}^\epsilon f(X^\epsilon_s, Z^\epsilon_s)\,ds\right.$$
$$\left. - \int_u^v \sum_j \partial_j f(X^\epsilon_s, Z^\epsilon_s)B^{ji}(X^\epsilon_s)\,ds \,\middle|\, \mathcal{H}^\epsilon_u\right] = 0,$$

hence the process in (C.1.2) is an $\{\mathcal{H}^\epsilon_t\}$-martingale. □

Proof of Lemma 5.2.82. Fix some $\phi \in C_c^\infty(\mathbb{R}^d)$, $\epsilon \in (0,1]$. Since $1 \in \mathcal{D}(\mathcal{Q})$ with $\mathcal{Q}1 \equiv 0$ (see Condition 4.1.52), by (4.2.19) we have $f := \phi \otimes 1 \in \tilde{\mathcal{D}}$. Thus, by (5.2.16) and (5.2.19) the process

$$\pi^\epsilon_t\phi - \int_0^t \mu^\epsilon_u\left(\mathcal{C}^\epsilon(\phi\otimes 1)\right)du, \quad t\in[0,T]$$

is a continuous $\{\mathcal{F}^{Y^\epsilon}_{t+}\}$-martingale, and so the sample paths of $\{\pi^\epsilon_t\phi,\ t\in[0,T]\}$ are continuous. Since by Fact D.1.129 the set $C_c^\infty(\mathbb{R}^d)$ is convergence determining, it follows that $\{\pi^\epsilon_t\}$ is a continuous process. Therefore, for every $t \in (0,T]$

$$\pi^\epsilon_t = \lim_{s\to t-}\pi^\epsilon_s,$$

and so, as every π^ϵ_s is $\mathcal{F}^{Y^\epsilon}_t$-measurable for $s<t$, the process $\{\pi^\epsilon_t\}$ is $\{\mathcal{F}^{Y^\epsilon}_t\}$-adapted. The result for $\{\bar{\pi}_t\}$ follows by an analogous argument. □

Proof of Lemma 5.2.85. Fix $t \in [0,T]$ and $\Phi \in \mathcal{D}(\mathbb{H}(\mathcal{G},\mathcal{H},c))$ of the form given in (5.2.22). Using Itô's formula from (5.2.20) we get

$$\Phi(\nu_t) = \Phi(\nu_0) + \sum_{i=1}^{n} \int_0^t \partial_i H(\nu_u\phi_1, \dots, \nu_u\phi_n)\, d(\nu_u\phi_i)$$
$$+ \frac{1}{2}\sum_{i,j=1}^{n} \int_0^t \partial_i\partial_j H(\nu_u\phi_1, \dots, \nu_u\phi_n)\, d\langle \nu_{\cdot}\phi_i, \nu_{\cdot}\phi_j\rangle_u. \quad \text{(C.1.13)}$$

From (5.2.20) we also have

$$\langle \nu_{\cdot}\phi_i, \nu_{\cdot}\phi_j\rangle_t = \sum_{k=1}^{r} \int_0^t R_{\mathcal{H}_k}(\phi_i, c^k, \nu_u) R_{\mathcal{H}_k}(\phi_j, c^k, \nu_u)\, du, \quad 1 \le i, j \le n,$$

so from (5.2.20) and (C.1.13)

$$\begin{aligned}\Phi(\nu_t) =&\Phi(\nu_0) + \sum_{i=1}^{n} \int_0^t \partial_i H(\nu_u\phi_1, \dots, \nu_u\phi_n)\nu_u(\mathcal{G}\phi_i)\, du \\ &+\frac{1}{2}\sum_{k=1}^{r}\sum_{i,j=1}^{n} \int_0^t \partial_i\partial_j H(\nu_u\phi_1, \dots, \nu_u\phi_n) R_{\mathcal{H}_k}(\phi_i, c^k, \nu_u) R_{\mathcal{H}_k}(\phi_j, c^k, \nu_u)\, du \\ &+\sum_{i=1}^{n}\sum_{k=1}^{r} \int_0^t \partial_i H(\nu_u\phi_1, \dots, \nu_u\phi_n) R_{\mathcal{H}_k}(\phi_i, c^k, \nu_u)\, dV_u^k. \qquad \text{(C.1.14)}\end{aligned}$$

Since

$$\mathrm{E}\left[\int_0^T |R_{\mathcal{H}_k}(\phi_i, c^k, \nu_u)|^2\, du\right] < \infty, \quad 1 \le i \le n,\ 1 \le k \le r,$$

and the second partial derivatives of H are uniformly bounded, the stochastic integral in (C.1.14) is an $\{\mathcal{F}_t\}$-martingale, whence the result follows. $\square$

Proof of Lemma 5.2.86. From (5.2.24) we have $\mathcal{D}(\mathbb{H}(\mathcal{G},\mathcal{H},c)) \subset \bar{C}(\mathcal{P}(E))$. To show that it is an algebra, fix $\Phi_i \in \mathcal{D}(\mathbb{H}(\mathcal{G},\mathcal{H},c))$, $i = 1, 2$, so that for some $(\phi_j^{(i)})_{j=1}^{n(i)} \subset \mathcal{D}(\mathcal{G},\mathcal{H})$ we have

$$\Phi_i(\nu) = H_i(\nu\phi_1^{(i)}, \nu\phi_2^{(i)}, \dots, \nu\phi_{n(i)}^{(i)}), \quad \nu \in \mathcal{P}(E),\ i = 1, 2.$$

Define

$$H := H_1 \otimes H_2.$$

Then $H \in C_c^\infty(\mathbb{R}^{n(1)+n(2)})$, and

$$\Phi_1(\nu)\Phi_2(\nu) = H(\nu\phi_1^{(1)}, \nu\phi_2^{(1)}, \dots, \nu\phi_{n(1)}^{(1)}, \nu\phi_1^{(2)}, \dots, \nu\phi_{n(2)}^{(2)}),$$

so we have that $\Phi_1\Phi_2 \in \mathcal{D}(\mathbb{H}(\mathcal{G}, \mathcal{H}, c))$.

Next we show that $\Phi_1 + \Phi_2 \in \mathcal{D}(\mathbb{H}(\mathcal{G}, \mathcal{H}, c))$. Since $(\phi_j^{(i)})_{j=1}^{n(i)} \subset \bar{C}(E)$, $i = 1, 2$, there exists $R \in [0, \infty)$ such that

$$|\nu\phi_j^{(i)}| \leq R, \quad \forall \nu \in \mathcal{P}(E),\ 1 \leq j \leq n(i),\ i = 1, 2. \tag{C.1.15}$$

Fix some $G \in C_c^\infty(\mathbb{R}^{n(1)+n(2)})$ such that $0 \leq G \leq 1$, and $G(x) = 1$ for every $x \in \mathbb{R}^{n(1)+n(2)}$ such that $|x^i| \leq R$, $1 \leq i \leq n(1) + n(2)$. Define

$$\begin{aligned} &H(x_1, x_2, \dots, x_{n(1)}, x_{n(1)+1}, \dots, x_{n(1)+n(2)}) := \\ &\quad G(x_1, x_2, \dots, x_{n(1)+n(2)})[H_1(x_1, x_2, \dots, x_{n(1)}) + H_2(x_{n(1)+1}, \dots, x_{n(1)+n(2)})], \end{aligned} \tag{C.1.16}$$

for every $x_j \in \mathbb{R}$. By (C.1.15) and the definition of G we have

$$G(\nu\phi_1^{(1)}, \dots, \nu\phi_{n(1)}^{(1)}, \nu\phi_1^{(2)}, \dots, \nu\phi_{n(2)}^{(2)}) = 1, \quad \forall \nu \in \mathcal{P}(E), \tag{C.1.17}$$

so from (C.1.16) it follows that

$$\Phi_1(\nu) + \Phi_2(\nu) = H(\nu\phi_1^{(1)}, \nu\phi_2^{(1)}, \dots, \nu\phi_{n(1)}^{(1)}, \nu\phi_1^{(2)}, \dots, \nu\phi_{n(2)}^{(2)}), \quad \forall \nu \in \mathcal{P}(E).$$

Thus, since (C.1.16) ensures $H \in C_c^\infty(\mathbb{R}^{n(1)+n(2)})$, we have $\Phi_1 + \Phi_2 \in \mathcal{D}(\mathbb{H}(\mathcal{G}, \mathcal{H}, c))$. Also, $\alpha\Phi_1 \in \mathcal{D}(\mathbb{H}(\mathcal{G}, \mathcal{H}, c))$, $\forall \alpha \in \mathbb{R}$, and so we have that $\mathcal{D}(\mathbb{H}(\mathcal{G}, \mathcal{H}, c))$ is an algebra. Furthermore, from (C.1.17) it follows that $\mathcal{D}(\mathbb{H}(\mathcal{G}, \mathcal{H}, c))$ contains constant functions.

To show that $\mathcal{D}(\mathbb{H}(\mathcal{G}, \mathcal{H}, c))$ separates points in $\mathcal{P}(E)$ if $\mathcal{D}(\mathcal{G}, \mathcal{H})$ is separating, fix some $\nu_1, \nu_2 \in \mathcal{P}(E)$, $\nu_1 \neq \nu_2$, and find $\phi \in \mathcal{D}(\mathcal{G}, \mathcal{H})$ such that $\nu_1\phi \neq \nu_2\phi$. Also, fix an $H \in C_c^\infty(\mathbb{R})$ such that $H(\nu_1\phi) \neq H(\nu_2\phi)$. Then the mapping

$$\Phi(\nu) := H(\nu\phi), \quad \nu \in \mathcal{P}(E)$$

belongs to $\mathcal{D}(\mathbb{H}(\mathcal{G},\mathcal{H},c))$, and $\Phi(\nu_1) \neq \Phi(\nu_2)$.

Similarly, to show that $\mathcal{D}(\mathbb{H}(\mathcal{G},\mathcal{H},c))$is strongly separates points in $\mathcal{P}(E)$ if $\mathcal{D}(\mathcal{G},\mathcal{H})$ is convergence determining, suppose that $(\nu_i)_{i=0}^{\infty} \subset \mathcal{P}(E)$ is such that for every $H \in C_c^{\infty}(\mathbb{R})$, $\phi \in \mathcal{D}(\mathcal{G},\mathcal{H})$ we have

$$\lim_{n\to\infty} H(\nu_n\phi) = H(\nu_0\phi).$$

Since $C_c^{\infty}(\mathbb{R})$ strongly separates points in $\mathbb{R}$, this implies

$$\lim_{n\to\infty} \nu_n\phi = \nu_0\phi,$$

which, in view of the fact that $\mathcal{D}(\mathcal{G},\mathcal{H})$ is convergence determining, implies that $(\nu_i)_{i=1}^{\infty}$ converges to ν_0 in $\mathcal{P}(E)$, completing the proof. □

Proof of Corollary 5.2.87. Since E is compact, $\mathcal{P}(E)$ is compact by Prohorov theorem (see Theorem 3.2.2 of Ethier and Kurtz [8]). By Lemma 5.2.86 the collection $\mathcal{D}(\mathbb{H}(\mathcal{G},\mathcal{H},c))$ is an algebra that separates points, so the result follows by the Stone-Weierstrass theorem. □

Proof of Lemma 5.2.89. Fix some $t \in [0,T]$, $g \in C_c^{\infty}(\mathbb{R}^r) \cup \{1\}$, and let $\Phi \in \mathcal{D}(\mathbb{H}(\mathcal{G},\mathcal{H},c))$ be of the form given in (5.2.22). Using the fact that $\{V_t\}$ is an $\{\mathcal{F}_t\}$-Wiener process, by Itô's formula we get

$$g(V_t) = g(0) + \frac{1}{2}\int_0^t \Delta g(V_u)\,du + \int_0^t \sum_{k=1}^{r} \partial_k g(V_u)\,dV_u^k.$$

Thus, by the Itô's product rule and (C.1.14) we have

$$\begin{aligned}\Phi(\nu_t)g(V_t) &= \Phi(\nu_0)g(0) + \frac{1}{2}\int_0^t \Phi(\nu_u)\Delta g(V_u)\,du \\ &+ \sum_{k=1}^r \int_0^t \Phi(\nu_u)\partial_k g(V_u)\,dV_u^k + \int_0^t g(V_u)\mathbb{H}(\mathcal{G},\mathcal{H},c)(\Phi)(\nu_u)\,du \\ &+ \sum_{k=1}^r \sum_{i=1}^n \int_0^t \partial_i H(\nu_u\phi_1,\dots,\nu_u\phi_n)R_{\mathcal{H}_k}(\phi_i,c^k,\nu_u)\partial_k g(V_u)\,du \\ &+ \sum_{k=1}^r \sum_{i=1}^n g(V_u)\partial_i H(\nu_u\phi_1,\dots,\nu_u\phi_n)R_{\mathcal{H}_k}(\phi_i,c^k,\nu_u)\,dV_u^k, \quad t\in[0,T].\end{aligned}$$

Since the two stochastic integrals on the right-hand side are $\{\mathcal{F}_t\}$-martingales (recall that $g \in C_c^\infty(\mathbb{R}^r) \cup \{1\}$, $H \in C_c^\infty(\mathbb{R}^n)$), the result follows. □

C.2 Appendix for Section 5.3

For the proofs in this Appendix we will frequently need the following useful result, which is essentially given in Remark 2.5 of Kurtz [21]:

Lemma C.2.110. *For every $D \in \mathbb{N}$ there exists a countable set $\mathcal{H}_D \subset C_c^\infty(\mathbb{R}^D)$ with the following property: for each $f \in C_c^\infty(\mathbb{R}^D)$ there exists a sequence $\{f_n\} \subset \mathcal{H}_D$ such that for every $g : \mathbb{R}^d \to \mathbb{R}$ that is bounded on bounded sets we have*

$$\begin{aligned}&\lim_{n\to\infty}\|gf_n - gf\| = 0,\\ &\lim_{n\to\infty}\|g\,\partial_i f_n - g\,\partial_i f\| = 0, \quad 1\le i\le D\\ &\lim_{n\to\infty}\|g\,\partial_i\partial_j f_n - g\,\partial_i\partial_j f\| = 0, \quad 1\le i,j\le D.\end{aligned}$$

Proof. Without loss of generality we may suppose that $g \equiv 1$, as

$$\begin{aligned}&\|gf_n - gf\| \le \|gI_A\|\,\|f_n - f\|,\\ &\|g\,\partial_i f_n - g\,\partial_i f\| \le \|gI_A\|\,\|\partial_i f_n - \partial_i f\|, \quad 1\le i\le D\\ &\|g\,\partial_i\partial_j f_n - g\,\partial_i\partial_j f\| \le \|gI_A\|\,\|\partial_i\partial_j f_n - \partial_i\partial_j f\|, \quad 1\le i,j\le D,\end{aligned}$$

where $A \subset \mathbb{R}^D$ is the support of f. The key observation is the following: since

$$\mathcal{G}_D := \{(f, \partial_1 f, \partial_2 f, \dots, \partial_D f, \partial_1\partial_1 f, \partial_1\partial_2 f, \dots, \partial_D\partial_D f) : f \in C_c^\infty(\mathbb{R}^D)\}$$

is a subset of the separable set $\hat{C}(\mathbb{R}^D)^{1+D+D^2}$, it is also separable (under the supremum norm). Therefore, there exists a countable dense subset $\mathcal{H}'_D \subset \mathcal{G}_D$. If we now take $\mathcal{H}_D$ to be the collection of all first co-ordinates of $1+D+D^2$-tuples in $\mathcal{H}'_D$, it is immediate that $\mathcal{H}_D$ has the desired properties. □

Proof of Lemma 5.3.95. Fix $n \in \mathbb{N}$, and let $\mathcal{H}_n$, $\mathcal{H}_d$, and $\mathcal{H}_r$ be given by Lemma C.2.110. Define a countable set $\mathcal{O}_n \subset \mathcal{D}(\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h))$ by

$$\begin{aligned}\mathcal{O}_n := \{\Psi \in \bar{C}(\mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^r) : \Psi = \Phi \otimes g, \quad \text{where} \quad \Phi(\nu) = H(\nu\phi_1, \dots, \nu\phi_n),\\ \forall \nu \in \mathcal{P}(\mathbb{R}^d) \quad \text{for some} \quad H \in \mathcal{H}_n,\ \phi_i \in \mathcal{H}_d,\ 1 \le i \le n, \quad \text{and} \quad g \in \mathcal{H}_r\}. \end{aligned} \tag{C.2.18}$$

Now fix some $\Psi \in \mathcal{D}(\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h))$ given by

$$\Phi(\nu) := H(\nu\phi_1, \nu\phi_2, \dots, \nu\phi_n) g(y), \quad (\nu, y) \in \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^r, \tag{C.2.19}$$

for some $H \in C_c^\infty(\mathbb{R}^n)$, $\phi_i \in C_c^\infty(\mathbb{R}^d)$, $1 \le i \le n$, $g \in C_c^\infty(\mathbb{R}^r)$.

From Lemma C.2.110 there exist sequences $\{H_m\} \subset \mathcal{H}_n$, $\{\phi_{i,m}\} \subset \mathcal{H}_d$, $1 \le i \le n$, and $\{g_m\} \subset \mathcal{H}_r$ such that

$$\lim_{m\to\infty} \|H_m - H\| = 0, \tag{C.2.20a}$$

$$\lim_{m\to\infty} \|\partial_i H_m - \partial_i H\| = 0, \quad 1 \le i \le n, \tag{C.2.20b}$$

$$\lim_{m\to\infty} \|\partial_i\partial_j H_m - \partial_i\partial_j H\| = 0, \quad 1 \le i, j \le n \tag{C.2.20c}$$

$$\lim_{m\to\infty} \|\phi_{i,m} - \phi_i\| = 0, \quad 1 \le i \le n, \tag{C.2.20d}$$

$$\lim_{m\to\infty} \|\mathcal{L}\phi_{i,m} - \mathcal{L}\phi_i\| = 0, \quad 1 \le i \le n, \tag{C.2.20e}$$

$$\lim_{m\to\infty} \|\mathcal{B}_k\phi_{i,m} - \mathcal{B}_k\phi_i\| = 0, \quad 1 \le k \le r,\ 1 \le i \le n, \tag{C.2.20f}$$

$$\lim_{m\to\infty} \|g_m - g\| = 0. \tag{C.2.20g}$$

Put

$$\Psi_m(\nu) := H_m(\nu\phi_{1,m}, \dots, \nu\phi_{n,m})\, g_m(y), \quad \forall (\nu, y) \in \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^r,\ m \in \mathbb{N}. \quad (C.2.21)$$

Then, in view of the uniform convergence in (C.2.20) and the structure of $\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h)$ (recall (5.2.25b)), we have

$$\lim_{m\to\infty} \|\Psi_m - \Psi\| = 0, \quad \lim_{m\to\infty} \|\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h)(\Psi_n) - \hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h)(\Psi)\| = 0. \quad (C.2.22)$$

Now put

$$\mathcal{O} := \left\{ q + \Psi : \Psi \in \bigcup_n \mathcal{O}_n,\ q \in \mathbb{Q} \right\}. \quad (C.2.23)$$

Then from (C.2.22) it follows that $\mathcal{O}$ has the desired properties. □

Proof of Lemma 5.3.97. Fix $t \in [0, T]$. By Theorem 4.2.63 $\{X_t^{\epsilon_n}\}$ converges weakly to $\{\bar{X}_t\}$ as $n \to \infty$, hence the collection $(\mathcal{L}(X_t^{\epsilon_n}),\ n \in \mathbb{N})$ is relatively compact in $\mathcal{P}(\mathbb{R}^d)$. Thus, for every $\delta > 0$ there exists a compact set $K_\delta \subset \mathbb{R}^d$ such that

$$P(X_t^{\epsilon_n} \in K_\delta) \geq 1 - \delta, \quad \forall n \in \mathbb{N}. \quad (C.2.24)$$

Let $a \in (0, 1)$ be arbitrary. Since for every $n \in \mathbb{N}$

$$\begin{aligned} P(X_t^{\epsilon_n} \in K_\delta) &= \mathrm{E}[I_{K_\delta}(X_t^{\epsilon_n})] \\ &= \mathrm{E}[\mathrm{E}[I_{K_\delta}(X_t^{\epsilon_n}) | \mathcal{F}_t^{Y^{\epsilon_n}}]] \\ &= \mathrm{E}[\pi_t^{\epsilon_n}(K_\delta)] \\ &= 1 - a + \mathrm{E}[\pi_t^{\epsilon_n}(K_\delta) - 1 + a] \\ &\leq 1 - a + \mathrm{E}[(\pi_t^{\epsilon_n}(K_\delta) - 1 + a) I_{\{\pi_t^{\epsilon_n}(K_\delta) - 1 + a > 0\}}] \\ &\leq 1 - a + aP(\pi_t^{\epsilon_n}(K_\delta) > 1 - a), \end{aligned}$$

by (C.2.24) it follows that

$$1 - \delta < 1 - a + aP(\{\pi_t^{\epsilon_n}(K_\delta) > 1 - a\}), \quad \forall n \in \mathbb{N}. \quad (C.2.25)$$

Fix some $\eta \in (0,\infty)$. For every $k \in \mathbb{N}$ let $a_k := 1/k$, and let $K_k^\eta \subset \mathbb{R}^d$ be the compact set corresponding to $\delta_k := \eta/(k2^k)$ in (C.2.24). Then by (C.2.25) we get

$$P(\pi_t^{\epsilon_n}(K_k^\eta) > 1 - 1/k) \geq 1 - \eta/2^k, \quad \forall k \in \mathbb{N},\ n \in \mathbb{N},$$

which implies

$$P\left(\bigcap_{k=1}^{\infty}\{\pi_t^{\epsilon_n}(K_k^\eta) > 1 - 1/k\}\right) \geq 1 - \eta, \quad \forall n \in \mathbb{N}. \tag{C.2.26}$$

Define $\mathcal{K}^\eta \subset \mathcal{P}(\mathbb{R}^d)$ as follows:

$$\mathcal{K}^\eta := \left\{\mu \in \mathcal{P}(\mathbb{R}^d) : \mu(K_k^\eta) > 1 - 1/k, \quad \forall k \in \mathbb{N}\right\}.$$

Since the set $\mathcal{K}^\eta$ is tight by definition, by the Prohorov theorem (see Theorem 3.2.2 of Ethier and Kurtz [8]), its closure $\bar{\mathcal{K}}^\eta$ is compact in the topology of $\mathcal{P}(\mathbb{R}^d)$. Therefore, by (C.2.26) and the definition of $\mathcal{K}^\eta$ we have

$$P(\pi_t^{\epsilon_n} \in \bar{\mathcal{K}}^\eta) \geq 1 - \eta, \quad \forall n \in \mathbb{N},$$

so it follows that the collection $(\mathcal{L}(\pi_t^{\epsilon_n}),\ n \in \mathbb{N})$ is relatively compact in $\mathcal{P}(\mathcal{P}(\mathbb{R}^d))$.

□

Proof of Fact 5.3.99. (i) Since $H \in C_c^\infty(\mathbb{N}^m)$ (see (5.3.33)) it is Lipshitz continuous, so there exists $K \in [0,\infty)$ such that for every $n \in \mathbb{N}, t \in [0,T]$ (recall (5.2.15))

$$\begin{aligned}
|H(\mu_t^n f^{n,\phi_1}, \dots, \mu_t^n f^{n,\phi_m}) - H(\pi_t^n \phi_1, \dots, \pi_t^n \phi_m)| \\
&\leq K\,\epsilon_n \sum_{i=1}^{m} |\mu_t^n(f_1^{\phi_i} + \epsilon f_2^{\phi_i} + \epsilon f_3^{\phi_i})| \\
&\leq K\,\epsilon_n \sum_{i=1}^{m} (\|f_1^{\phi_i}\| + \epsilon\|f_2^{\phi_i}\| + \epsilon\|f_3^{\phi_i}\|),
\end{aligned}$$

as required.

(ii) This follows directly from (5.3.36b) and Remark 5.2.88.

(iii) This part follows by the same argument as (i); since the calculations are straightforward, the details are omitted. □

C.3 Appendix for Theorem 5.3.96

The goal of this appendix is to prove Theorem 5.3.96.

Remark C.3.111. Fix some arbitrary $x = \mu \times y$, $\mu \in \mathcal{P}(\mathbb{R}^d)$, $y \in \mathbb{R}^r$. If we replace x_0 in the initial condition (4.4.41) by some random variable X_0 with $\mathcal{L}(X_0) = \mu$, i.e. if we consider

$$d\bar{X}_t^i = b^i(\bar{X}_t)\,dt + \sum_{j=1}^{d} \sigma^{ij}(\bar{X}_t)\,d\bar{V}_t^j + \sum_{j=1}^{r} B^{ij}(\bar{X}_t)\,d\bar{W}_t^j, \quad 1 \le i \le d, \quad t \in [0,T],$$

with

$$\bar{X}_0 = X_0,$$

then the resulting weak solution $\{(\bar{\Omega}, \bar{\mathcal{F}}, \{\bar{\mathcal{F}}_t\}, \bar{P}), (\bar{X}_t, \hat{W}_t)\}$ together with the observation process $\{\bar{Y}_t\}$ (see (5.1.7)) gives rise to the nonlinear filter $\{\bar{\pi}_t\}$ for which $\mathcal{L}(\bar{\pi}_0) = \delta_\mu$. Thus, with $\hat{I}_t \equiv \bar{I}_t + y$, the pair $\{(\bar{\Omega}, \bar{\mathcal{F}}, \{\bar{\mathcal{F}}_t\}, \bar{P}), (\bar{\pi}_t, \hat{I})\}$ is a solution of the martingale problem for $\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h)$ with $\mathcal{L}((\bar{\pi}_0, \hat{I}_0)) = \delta_x$.

Proof of Theorem 5.3.96. Since by Theorem C.3.112 every solution of the martingale problem for $\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h)$ has continuous (hence corlol) modification, it suffices to establish well-posedness of the *corlol* martingale problem for $\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h)$. Next, in view of Remark 5.2.88 and the fact that Condition (I) of Theorem 5.3.91 has been verified, Theorem D.2.138 shows that we only have to establish well posedness of the martingale problem for $(\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h), \delta_x)$ for each $x \in \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^r$.

Fix $x \in \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^r$. Uniqueness for the martingale problem for $(\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h), \delta_x)$ follows from Theorem C.3.112 and Theorem 3.3.34(ii), while existence follows from Remark C.3.111. $\square$

Theorem C.3.112. *Suppose that* $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{P}), (\tilde{\nu}_t, \tilde{V}_t)\}$ *is a (progressively measurable) solution of the martingale problem for* $\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h)$. *Then there exists some*

$\mathcal{P}(\mathbb{R}^d)\times\mathbb{R}^r$*-valued, continuous, and* $\{\tilde{\mathcal{F}}_t\}$*-adapted process* $\{(\tilde{\nu}'_t, \tilde{V}'_t)\}$ *on* $(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\mathcal{F}_t\}, \tilde{P})$ *such that*

(a) $\{(\tilde{V}'_t - \tilde{V}'_0, \tilde{\mathcal{F}}_t)\}$ *is an* $\mathbb{R}^r$*-valued Wiener process;*

(b) $\{(\tilde{\nu}'_t, \tilde{V}'_t)\}$ *is a modification of* $\{(\tilde{\nu}_t, \tilde{V}_t)\}$*;*

(c) for each $\phi \in C_c^\infty(\mathbb{R}^d)$ *we have a.s.*

$$\tilde{\nu}'_t\phi = \tilde{\nu}'_0\phi + \int_0^t \tilde{\nu}'_s(\mathcal{L}\phi)\, ds + \sum_{k=1}^r \int_0^t R_{\mathcal{B}_k}(\phi, h^k, \tilde{\nu}'_s)\, d(\tilde{V}'_s)^k, \quad \forall t \in [0,T]. \quad \text{(C.3.27)}$$

Proof. Fix some $\phi \in C_c^\infty(\mathbb{R}^d)$, and find $H_1 \in C_c^\infty(\mathbb{R})$ such that

$$H_1(x) = 1, \qquad \forall\, x \in \mathbb{R}, \quad \text{such that} \quad |x| \le \|\phi\|. \quad \text{(C.3.28)}$$

Define $\Phi_1 \in \mathcal{D}(\mathbb{H}(\mathcal{L}, \mathcal{B}, h))$ by

$$\Phi_1(\nu) := H_1(\nu\phi), \quad \forall\, \nu \in \mathcal{P}(\mathbb{R}^d). \quad \text{(C.3.29)}$$

It follows from (C.3.28) and (C.3.29) that

$$\Phi_1(\nu) = 1, \quad \forall\, \nu \in \mathcal{P}(\mathbb{R}^d). \quad \text{(C.3.30)}$$

Since

$$\partial H_1(\nu\phi) = 0 \quad \text{and} \quad \partial^2 H_1(\nu\phi) = 0, \quad \forall\, \nu \in \mathcal{P}(\mathbb{R}^d), \quad \text{(C.3.31)}$$

we have

$$\mathbb{H}(\mathcal{L}, \mathcal{B}, h)(\Phi_1)(\nu) = 0, \qquad \forall\, \nu \in \mathcal{P}(\mathbb{R}^d). \quad \text{(C.3.32)}$$

Now fix some $g \in \{1\} \cup C_c^\infty(\mathbb{R}^r)$ and define $\Psi_1 \in \mathcal{D}(\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h))$ by

$$\Psi_1 := \Phi_1 \otimes g. \quad \text{(C.3.33)}$$

From (C.3.30) and (C.3.33):

$$\Psi_1(\nu, y) = g(y), \quad \forall\ (\nu, y) \in \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^r, \tag{C.3.34}$$

while from (C.3.30), (C.3.31), (C.3.32), and (C.3.33), we have

$$\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h)(\Psi_1)(\nu, y) = \frac{1}{2}\Delta g(y) \quad \forall\ (\nu, y) \in \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^r, \tag{C.3.35}$$

hence (C.3.34) and (C.3.35) show that

$$g(\tilde{V}_t) - \frac{1}{2}\int_0^t \Delta g(\tilde{V}_s)\, ds, \quad t \in [0, T]$$

is a $\{\tilde{\mathcal{F}}_t\}$- martingale for each $g \in C_c^\infty(\mathbb{R}^r)$. Thus, Proposition 5.3.5 of Ethier and Kurtz [8] ensures that $\{\tilde{V}_t\}$ has a *continuous* modification $\{\tilde{V}_t'\}$, which is $\{\tilde{\mathcal{F}}_t\}$-adapted, and Theorem 4.1.1 of Stroock and Varadhan [35] shows that $\{(\tilde{V}_t' - \tilde{V}_0', \tilde{\mathcal{F}}_t)\}$ is an $\mathbb{R}^r$-valued Wiener process.

Fix some $\Phi \in \mathcal{D}(\mathbb{H}(\mathcal{L}, \mathcal{B}, h))$ and put

$$\Psi := \Phi \otimes 1. \tag{C.3.36}$$

Then

$$\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h)(\Psi)(\nu, y) = \mathbb{H}(\mathcal{L}, \mathcal{B}, h)(\Phi)(\nu), \quad \forall \nu \in \mathcal{P}(\mathbb{R}^d),\ y \in \mathbb{R}^r. \tag{C.3.37}$$

By (C.3.36) and (C.3.37)

$$\Phi(\tilde{\nu}_t) - \int_0^t \mathbb{H}(\mathcal{L}, \mathcal{B}, h)(\Phi)(\tilde{\nu}_s)\, ds, \quad t \in [0, T] \tag{C.3.38}$$

is an $\{\tilde{\mathcal{F}}_t\}$-martingale for every $\Phi \in \mathcal{D}(\mathbb{H}(\mathcal{L}, \mathcal{B}, h))$. We now show that $\{\tilde{\nu}_t\}$ has a *corlol* modification. Since by Lemma C.3.114(ii) the set $\mathcal{P}(\mathbb{R}^d)$ is homeomorphic to the set $\{\nu \in \mathcal{P}(\mathbb{R}^{d*}) : \nu(\Delta) = 0\}$, to accomplish this task it suffices to show that $\{\tilde{\nu}_t\}$, regarded as a $\mathcal{P}(\mathbb{R}^{d*})$-valued process, has a ($\mathcal{P}(\mathbb{R}^{d*})$-valued) corlol modification $\{\tilde{\nu}_t'\}$ such that

$$\tilde{P}\big(\tilde{\nu}_t'(\mathbb{R}^d) = 1,\ \forall t \in [0, T]\big) = 1. \tag{C.3.39}$$

To that end, note that, since

$$\tilde{\nu}_t(\Delta) = 0, \quad \forall t \in [0,T]$$

from (C.3.38) it follows that the process

$$\Phi(\tilde{\nu}_t) - \int_0^t \mathbb{H}(\mathcal{L}^\Delta, \mathcal{B}^\Delta, h^\Delta)(\Phi)(\tilde{\nu}_s)\, ds, \quad t \in [0,T]$$

is an $\{\tilde{\mathcal{F}}_t\}$-martingale for every $\Phi \in \mathcal{D}(\mathbb{H}(\mathcal{L}^\Delta, \mathcal{B}^\Delta, h^\Delta))$. Now

$$\mathcal{D}(\mathcal{L}^\Delta, \mathcal{B}^\Delta) := \{\phi \in \bar{C}(\mathbb{R}^{d*}) : \phi|_{\mathbb{R}^d} \in C_c^\infty(\mathbb{R}^d),\ \phi(\Delta) = 0\},$$

so it separates probability measures on $\mathcal{P}(\mathbb{R}^{d*})$. Therefore, by Corollary 5.2.87 the collection $\mathbb{H}(\mathcal{L}^\Delta, \mathcal{B}^\Delta, h^\Delta)$ is dense in $\bar{C}(\mathcal{P}(\mathbb{R}^{d*}))$, which, due to compactness of $\mathcal{P}(\mathbb{R}^{d*})$, implies that $\mathbb{H}(\mathcal{L}^\Delta, \mathcal{B}^\Delta, h^\Delta)$ has a countable subset that separates points. Hence, by Theorem 4.3.6 of Ethier and Kurtz [8] there exists a *corlol* process $\{\tilde{\nu}'_t\}$ which is a modification of $\{\tilde{\nu}_t\}$. Since $\tilde{\mathcal{F}}_0$ includes all $\tilde{P}$-null events in $\tilde{\mathcal{F}}$, one sees that $\{\tilde{\nu}'_t\}$ is also $\{\tilde{\mathcal{F}}_t\}$-adapted, so we have that

$$\Phi(\tilde{\nu}'_t) - \int_0^t \mathbb{H}(\mathcal{L}^\Delta, \mathcal{B}^\Delta, h^\Delta)(\Phi)(\tilde{\nu}'_u)\, du, \quad t \in [0,T]$$

is an $\{\tilde{\mathcal{F}}_t\}$-martingale for every $\Phi \in \mathcal{D}(\mathbb{H}(\mathcal{L}^\Delta, \mathcal{B}^\Delta, h^\Delta))$. Since $\tilde{\nu}_0 \equiv \tilde{\nu}'_0(\mathbb{R}^d) \equiv 1$, we have

$$\tilde{P}(\tilde{\nu}'_0(\mathbb{R}^d) = 1) = 1, \tag{C.3.40}$$

so from Lemma C.3.113 and (C.3.40) we get (C.3.39). Therefore, $\{\tilde{\nu}_t\}$ has a corlol modification $\{\tilde{\nu}'_t\}$. Since, in addition, $\{\tilde{V}'_t\}$ is a continuous modification of $\{\tilde{V}_t\}$, it follows from (C.3.38) that

$$\Psi(\tilde{\nu}'_t, \tilde{V}'_t) - \int_0^t \hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h)(\Psi)(\tilde{\nu}'_u, \tilde{V}'_u)\, du \quad t \in [0,T] \tag{C.3.41}$$

is an $\{\tilde{\mathcal{F}}_t\}$-martingale for every $\Psi \in \mathcal{D}(\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h))$.

It remains to show (C.3.27): fix $\phi \in C_c^\infty(\mathbb{R}^d)$ and put for $t \in [0,T]$

$$\eta_t := \tilde{\nu}'_t\phi - \int_0^t \tilde{\nu}'_u(\mathcal{L}\phi)\,du \tag{C.3.42}$$

$$\rho_t := \eta_t - \sum_{k=1}^r \int_0^t R_{\mathcal{B}_k}(\phi, h^k, \tilde{\nu}'_s)\,d(\tilde{V}'_s)^k. \tag{C.3.43}$$

Fix $H_2 \in C_c^\infty(\mathbb{R})$ such that

$$H_2(x) = x, \quad |x| \le \|\phi\|. \tag{C.3.44}$$

Define $\Phi_2 \in \mathcal{D}(\mathbb{H}(\mathcal{L}, \mathcal{B}, h))$ by

$$\Phi_2(\nu) := H_2(\nu\phi), \quad \forall \nu \in \mathcal{P}(\mathbb{R}^d). \tag{C.3.45}$$

From (C.3.44) we have

$$\Phi_2(\nu) = \nu\phi, \quad \forall \nu \in \mathcal{P}(\mathbb{R}^d).$$

Also, define $\Psi_2 \in \mathcal{D}(\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h))$ by

$$\Psi_2 := \Phi_2 \otimes 1, \tag{C.3.46}$$

so that

$$\Psi_2(\nu, y) = \nu\phi, \quad \forall \nu \in \mathcal{P}(\mathbb{R}^d),\ y \in \mathbb{R}^r.$$

Then from (C.3.44)

$$\mathbb{H}(\mathcal{L}, \mathcal{B}, h)(\Phi_2)(\nu) = \nu(\mathcal{L}\phi), \quad \nu \in \mathcal{P}(\mathbb{R}^d), \tag{C.3.47}$$

and from (C.3.46) and (C.3.47)

$$\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h)(\Psi_2)(\nu, y) = \mathbb{H}(\mathcal{L}, \mathcal{B}, h)(\Phi_2)(\nu) = \nu(\mathcal{L}\phi), \quad \forall \nu \in \mathcal{P}(\mathbb{R}^d). \tag{C.3.48}$$

From (C.3.41), (C.3.47), and (C.3.48) it follows that the process

$$\tilde{\nu}'_t\phi - \int_0^t \tilde{\nu}'_u(\mathcal{L}\phi)\,du, \quad t \in [0,T]$$

is an $\{\tilde{\mathcal{F}}_t\}$-martingale, so the same holds for the corlol process $\{\eta_t\}$ (see (C.3.42)). Also, since $\mathcal{D}(\mathbb{H}(\mathcal{L},\mathcal{B},h))$ is an algebra in $\bar{C}(\mathcal{P}(\mathbb{R}^d))$, we have $\Phi_2^2 \in \mathcal{D}(\mathbb{H}(\mathcal{L},\mathcal{B},h))$, hence

$$\Psi_2^2 \equiv \Phi_2^2 \otimes 1$$

is a member of $\mathcal{D}(\hat{\mathbb{H}}(\mathcal{L},\mathcal{B},h))$ with

$$\Psi_2^2(\nu,y) = H_2^2(\nu\phi), \quad \nu \in \mathcal{P}(\mathbb{R}^d),\ y \in \mathbb{R}^r. \tag{C.3.49}$$

Thus from (C.3.44) and (C.3.49) we have

$$\begin{aligned}\hat{\mathbb{H}}(\mathcal{L},\mathcal{B},h)(\Psi_2^2)(\nu,y) =& \partial(H_2^2)(\nu\phi)\nu(\mathcal{L}\phi) + \frac{1}{2}\partial^2(H_2^2)(\nu\phi)\sum_{k=1}^r R_{\mathcal{B}_k}^2(\phi,h^k,\nu) \\ =& 2(\nu\phi)\,\nu(\mathcal{L}\phi) + \sum_{k=1}^r R_{\mathcal{B}_k}^2(\phi,h^k,\nu). \end{aligned}\tag{C.3.50}$$

From (C.3.41),(C.3.49), (C.3.44), and (C.3.50), it follows that

$$(\tilde{\nu}_t'\phi)^2 - \int_0^t \left[2(\tilde{\nu}_s'\phi)\tilde{\nu}_s'(\mathcal{L}\phi) + \sum_{k=1}^r R_{\mathcal{B}_k}^2(\phi,h^k,\tilde{\nu}_s')\right] ds \tag{C.3.51}$$

is an $\{\tilde{\mathcal{F}}_t\}$-martingale. From (C.3.42) and (C.3.51), and Lemma D.3.145 we see that

$$\begin{aligned}&\eta_t^2 - \int_0^t \left[2(\tilde{\nu}_s'\phi)\tilde{\nu}_s'(\mathcal{L}\phi) + \sum_{k=1}^r R_{\mathcal{B}_k}^2(\phi,h^k,\tilde{\nu}_s') - 2(\tilde{\nu}_s'\phi)\tilde{\nu}_s'(\mathcal{L}\phi)\right] ds \\ &= \eta_t^2 - \int_0^t \sum_{k=1}^r R_{\mathcal{B}_k}^2(\phi,h^k,\tilde{\nu}_s')\, ds \end{aligned}\tag{C.3.52}$$

is an $\{\tilde{\mathcal{F}}_t\}$-martingale. Next, we write (recall (C.3.43))

$$\rho_t^2 = \eta_t^2 - 2\eta_t \sum_{k=1}^r \int_0^t R_{\mathcal{B}_k}(\phi,h^k,\tilde{\nu}_u')\, d(\tilde{V}_u')^k - \left(\sum_{k=1}^r \int_0^t R_{\mathcal{B}_k}(\phi,h^k,\tilde{\nu}_u')\, d(\tilde{V}_u')^k\right)^2,$$

so that

$$\mathrm{E}[\rho_t^2] = \mathrm{E}[\eta_t^2] - 2\sum_{k=1}^r \mathrm{E}\left[\eta_t \int_0^t R_{\mathcal{B}_k}(\phi,h^k,\tilde{\nu}_u')\, d(\tilde{V}_u')^k\right] + \sum_{k=1}^r \mathrm{E}\left[\int_0^t R_{\mathcal{B}_k}^2(\phi,h^k,\tilde{\nu}_u')\, du\right]. \tag{C.3.53}$$

Also since the process in (C.3.52) is an $\{\tilde{\mathcal{F}}_t\}$-martingale, we have

$$\mathrm{E}\left[\eta_t^2 - \int_0^t \sum_{k=1}^r R^2_{\mathcal{B}_k}(\phi, h^k, \tilde{\nu}'_u)\, du\right] = \mathrm{E}[\eta_0^2]. \tag{C.3.54}$$

By (C.3.53) and (C.3.54)

$$\mathrm{E}[\rho_t^2] = \mathrm{E}[\eta_0^2] - 2\sum_{k=1}^r \mathrm{E}\left[\eta_t \int_0^t R_{\mathcal{B}_k}(\phi, h^k, \tilde{\nu}'_u)\, d(\tilde{V}'_u)^k\right] + 2\sum_{k=1}^r \mathrm{E}\left[\int_0^t R^2_{\mathcal{B}_k}(\phi, h^k, \tilde{\nu}'_u)\, du\right]. \tag{C.3.55}$$

Since $\{\tilde{\nu}'_t\}$ is corlol, it follows that $\{R_k(t)\}$ defined by

$$R_k(t) := R_{\mathcal{B}_k}(\phi, h^k, \tilde{\nu}'_t), \quad t \in [0,T] \tag{C.3.56}$$

is corlol. Thus for every ω the set

$$\{t \in [0,T] : R_k(t) \neq R_k(t-)\}$$

is countable, hence has Lebesgue measure zero. Therefore

$$\mathrm{E}\left[\left(\int_0^t R_k(s)\, d(\tilde{V}'_s)^k - \int_0^t R_k(s-)\, d(\tilde{V}'_s)^k\right)^2\right] = \mathrm{E}\left[\int_0^t (R_k(s) - R_k(s-))^2\, ds\right] = 0,$$

thus

$$\tilde{P}\left(\int_0^t R_k(s)\, d(\tilde{V}'_s)^k = \int_0^t R_k(s-)\, d(\tilde{V}'_s)^k, \quad \forall t \in [0,T]\right) = 1. \tag{C.3.57}$$

Since $\{(\eta_t, \tilde{\mathcal{F}}_t)\}$ is a corlol martingale, and $\{(\tilde{V}'_t, \tilde{\mathcal{F}}_t)\}$ is a continuous martingale, Theorem VI.37.8 of Rogers and Williams [31] gives a unique continuous, $\{\tilde{\mathcal{F}}_t\}$-adapted process of finite variation, denoted by $[\eta, (\tilde{V}')^k]_t$, that is null at the origin and such that

$$\eta_t\, (\tilde{V}'_t)^k - [\eta, (\tilde{V}')^k]_t, \quad t \in [0,T]$$

is an $\{\tilde{\mathcal{F}}_t\}$-martingale. Since $\{R_k(t-)\}$ is left-continuous and $\{\tilde{\mathcal{F}}_t\}$-adapted, it is $\{\tilde{\mathcal{F}}_t\}$-predictable. Since it is also uniformly bounded on $[0,\infty) \times \tilde{\Omega}$, Theorem

VI.37.9(vi) from Rogers and Williams [31] implies that the process

$$(\eta_t - \eta_0) \int_0^t R_k(s-)\, d(\tilde{V}_s')^k - \int_0^t R_k(s-)\, d[\eta, (\tilde{V}')^k]_s$$

is an $\{\tilde{\mathcal{F}}_t\}$ martingale. Since $\{[\eta, (\tilde{V}')^k]_t\}$ is continuous, we have

$$\tilde{P}\left(\int_0^t R_k(s)\, d[\eta, (\tilde{V}')^k]_s = \int_0^t R_k(s-)\, d[\eta, (\tilde{V}')^k]_t, \quad \forall t \in [0, T]\right) = 1.$$

Thus, in view of (C.3.57), we see that

$$(\eta_t - \eta_0) \int_0^t R_k(s)\, (\tilde{V}_s')^k - \int_0^t R_k(s-)\, d[\eta, (\tilde{V}')^k]_s, \quad t \in [0, T]$$

is an $\{\tilde{\mathcal{F}}_t\}$ martingale. Since

$$\mathrm{E}\left[\eta_0 \int_0^t R_k(s)\, d(\tilde{V}_s')^k\right] = 0,$$

it follows that

$$\mathrm{E}\left[\eta_t \int_0^t R_k(s)\, d(\tilde{V}_s')^k\right] = \mathrm{E}\left[\int_0^t R_k(s)\, d[\eta, (\tilde{V}')^k]_s\right]. \tag{C.3.58}$$

From (C.3.55) and (C.3.58)

$$\mathrm{E}[\rho_t^2] = \mathrm{E}[\rho_0^2] - 2\sum_{k=1}^r \mathrm{E}\left[\int_0^t R_k(s)\, d[\eta, (\tilde{V}')^k]_s\right] + 2\sum_{k=1}^r \mathrm{E}\left[\int_0^t R_k^2(s)\, ds\right]. \tag{C.3.59}$$

We will show that

$$[\eta, (\tilde{V}')^k]_t = \int_0^t R_k(s)\, ds, \quad t \in [0, T]. \tag{C.3.60}$$

Assuming this is the case, from (C.3.59) and (C.3.60), we get

$$\mathrm{E}[\rho_t^2 - \rho_0^2] = 0,$$

hence, in view of the fact that $\{\rho_t\}$ is an $\{\tilde{\mathcal{F}}_t\}$-martingale, we get

$$\mathrm{E}[(\rho_t - \rho_0)^2] = 0,$$

hence

$$\rho_t = \rho_0 \quad a.s. \tag{C.3.61}$$

Since $\{\rho_t\}$ is corlol, this implies

$$\tilde{P}\big(\rho_t = \rho_0,\ \forall t \in [0,T]\big) = 1, \tag{C.3.62}$$

thus, from (C.3.42), (C.3.43), and (C.3.62) we get

$$\tilde{\nu}'_0\phi = \tilde{\nu}'_t\phi - \int_0^t \tilde{\nu}'_u(\mathcal{L}\phi)\,du - \sum_{k=1}^{r}\int_0^t R_{\mathcal{B}_k}(\phi, h^k, \tilde{\nu}'_s)\,d(\tilde{V}'_s)^k, \quad t \in [0,T],$$

which is (C.3.27). Since this holds for all $\phi \in C_c^\infty(\mathbb{R}^d)$, which is convergence determining, it follows that $\{\tilde{\nu}'_t\}$ is actually continuous (c.f. Remark 2.4.14).

To complete the proof, it remains to establish (C.3.60) For each $n = 1, 2, \ldots$ fix some $g_n \in C_c^\infty(\mathbb{R})$ with

$$g_n(y) = y^k, \quad \forall y \in \mathbb{R}^r,\ |y| \leq n \tag{C.3.63}$$

and put

$$\Psi_n := \Phi_2 \otimes g_n, \quad n = 3, 4, \ldots \tag{C.3.64}$$

Thus from (C.3.45)

$$\Psi_n(\nu, y) := H_2(\nu\phi)\,g_n(y), \quad \forall \nu \in \mathcal{P}(\mathbb{R}^d),\ y \in \mathbb{R}^r. \tag{C.3.65}$$

By (C.3.44) and (C.3.47) we have

$$\begin{aligned}\hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h)(\Psi_n)(\nu, y) &= y^k\,\mathbb{H}(\mathcal{L}, \mathcal{B}, h)(\Phi_2)(\nu) + \sum_{k=1}^{r}(\partial H_2)(\nu\phi)R_{\mathcal{B}_k}(\phi, h^k, \nu)\\ &= y^k\nu(\mathcal{L}\phi) + R_{\mathcal{B}_k}(\phi, h^k, \nu), \quad \forall |y| \leq n,\ \nu \in \mathcal{P}(\mathbb{R}^d),\ y \in \mathbb{R}^r.\end{aligned} \tag{C.3.66}$$

Also put

$$T_n := \inf\{t \in [0,T) : |(\tilde{V}'_t)^k| \geq n\}, \quad n \in \mathbb{N}. \tag{C.3.67}$$

Since T_n is an $\{\tilde{\mathcal{F}}_t\}$-stopping time, it follows from the optional stopping theorem (with $\Psi = \Psi_n$ in (C.3.41)) that the process

$$
\begin{aligned}
M_n(t) &:= \Psi_n(\tilde{\nu}'_{t\wedge T_n}, \tilde{V}'_{t\wedge T_n}) - \int_0^{t\wedge T_n} \hat{\mathbb{H}}(\mathcal{L}, \mathcal{B}, h)(\Psi_n)(\tilde{\nu}'_s, \tilde{V}'_s)\, ds \\
&= (\tilde{V}'_{t\wedge T_n})^k(\tilde{\nu}'_{t\wedge T_n}\phi) - \int_0^{t\wedge T_n} \left[(\tilde{V}'_s)^k\, \tilde{\nu}'_s(\mathcal{L}\phi) + R_{\mathcal{B}_k}(\phi, h^k, \tilde{\nu}'_s)\right] ds
\end{aligned} \tag{C.3.68}
$$

is an $\{\tilde{\mathcal{F}}_t\}$-martingale. From (C.3.68) and (C.3.42) we have

$$
\begin{aligned}
&\eta_{t\wedge T_n}(\tilde{V}'_{t\wedge T_n})^k - \int_0^{t\wedge T_n} R_{\mathcal{B}_k}(\phi, h^k, \tilde{\nu}'_s)\, ds \\
&= \left[(\tilde{\nu}'_{t\wedge T_n}\phi) - \int_0^{t\wedge T_n} \tilde{\nu}'_s(\mathcal{L}\phi)\, ds\right](\tilde{V}'_{t\wedge T_n})^k - \int_0^{t\wedge T_n} R_{\mathcal{B}_k}(\phi, h^k, \tilde{\nu}'_s)\, ds \\
&= M_n(t) + \left[\int_0^{t\wedge T_n} (\tilde{V}'_s)^k\, \tilde{\nu}'_s(\mathcal{L}\phi)\, ds - \left(\int_0^{t\wedge T_n} \tilde{\nu}'_s(\mathcal{L}\phi)\, ds\right)(\tilde{V}'_{t\wedge T_n})^k\right].
\end{aligned} \tag{C.3.69}
$$

By Lemma D.3.142 it follows that the process

$$
\int_0^t (\tilde{V}'_s)^k\, \tilde{\nu}'_s(\mathcal{L}\phi)\, ds - \left(\int_0^t \tilde{\nu}'_s(\mathcal{L}\phi)\, ds\right)(\tilde{V}'_t)^k, \quad t \in [0, T]
$$

is an $\{\tilde{\mathcal{F}}_t\}$-martingale, hence the second term on the RHS of (C.3.69) is a continuous $\{\tilde{\mathcal{F}}_t\}$-martingale. Since $\{M_n(t)\}$ is a corlol $\{\tilde{\mathcal{F}}_t\}$-martingale, we see from (C.3.69) that

$$
\eta_{t\wedge T_n}(\tilde{V}'_{t\wedge T_n})^k - \int_0^{t\wedge T_n} R_{\mathcal{B}_k}(\phi, h^k, \tilde{\nu}'_s)\, ds
$$

is a corlol $\{\tilde{\mathcal{F}}_t\}$ martingale for $n = 3, 4, \ldots$ Thus

$$
\eta_t(\tilde{V}'_t)^k - \int_0^t R_{\mathcal{B}_k}(\phi, h^k, \tilde{\nu}'_s)\, ds
$$

is a corlol $\{\tilde{\mathcal{F}}_t\}$-local martingale, whence (C.3.60) follows by the uniqueness of $[\eta, \tilde{V}']_t$. □

Lemma C.3.113. *Suppose that Condition* **AII** *holds (see Remark 5.1.76). Let* $(\hat{\Omega}, \hat{\mathcal{F}}, \{\hat{\mathcal{F}}_t\}, \hat{P})$ *be a complete filtered probability space on which is defined some*

$\mathcal{P}(\mathbb{R}^{d*})$-valued, corlol, $\{\hat{\mathcal{F}}_t\}$-adapted process $\{\hat{\nu}_t\}$ which solves the martingale problem for $\mathcal{H}(\mathcal{L}^\Delta, \mathcal{B}^\Delta, h^\Delta)$. If $\hat{P}(\hat{\nu}_0(\mathbb{R}^d) = 1) = 1$, then

$$\hat{P}(\hat{\nu}_t(\mathbb{R}^d) = 1,\ \forall t \in [0,T]) = 1. \tag{C.3.70}$$

Proof. Fix some $\phi \in C_c^\infty(\mathbb{R}^d)$ such that $\phi(x) = 1$ for $|x| \leq 1$, and for every $n \in \mathbb{N}$ let

$$\phi_n(x) = \begin{cases} \phi(x/n), \ x \in \mathbb{R}^d; \\ 0, \ x = \Delta. \end{cases}$$

Fix $H \in C_c^\infty(\mathbb{R})$ such that

$$H(x) = x, \quad |x| \leq 1, \tag{C.3.71}$$

and put

$$\Phi_n(\nu) := H(\nu\,\phi_n), \quad \forall \nu \in \mathcal{P}(\mathbb{R}^{d*}).$$

Since $\phi_n \in \mathcal{D}(\mathcal{L}^\Delta, \mathcal{B}^\Delta)$, we see that $\Phi_n \in \mathcal{D}(\mathcal{H}(\mathcal{L}^\Delta, \mathcal{B}^\Delta, h^\Delta))$, and

$$\mathcal{H}(\mathcal{L}^\Delta, \mathcal{B}^\Delta, h^\Delta)(\Phi_n)(\nu) = \nu(\mathcal{L}^\Delta \phi_n), \quad \nu \in \mathcal{P}(\mathbb{R}^{d*}). \tag{C.3.72}$$

By (C.3.72) the process

$$N_t^n := \hat{\nu}_t \phi_n - \int_0^t \hat{\nu}_s(\mathcal{L}^\Delta \phi_n)\, ds, \quad t \in [0,T]$$

is an $\{\hat{\mathcal{F}}_t^{\hat{\nu}}\}$-martingale. Since the paths of $\{\hat{\nu}_t\}$ are corlol, by the Lèvy's "backward" theorem, $\{N_t^n\}$ is also an $\{\hat{\mathcal{F}}_{t+}^{\hat{\nu}}\}$-martingale. Let

$$\begin{aligned} M_k &:= \{\mu \in \mathcal{P}(\mathbb{R}^{d*}) : \mu(\{\Delta\}) < 1/k\}, \quad k \in \mathbb{N}, \\ M &:= \bigcap_{k=1}^\infty M_k = \{\mu \in \mathcal{P}(\mathbb{R}^{d*}) : \mu(\{\Delta\}) = 0\}, \end{aligned} \tag{C.3.73}$$

and, with $d(\cdot,\cdot)$ denoting the Prohorov metric on $\mathcal{P}(\mathbb{R}^{d*})$ (see Remark D.1.116), define

$$T_m^k := \inf\left\{t \in [0,T) : \inf_{y\in\mathcal{P}(\mathbb{R}^{d*})\setminus M_k} d(y,\hat{\nu}_t) < \frac{1}{m}\right\}, \quad \forall\, k,m \in \mathbb{N}.$$

Since by Lemma C.3.114(i) for every $k \in \mathbb{N}$ the set M_k is open in $\mathcal{P}(\mathbb{R}^{d*})$, the sequence $(T_m^k)_{m\in\mathbb{N}}$ is a sequence of $\{\mathcal{F}_{t+}^{\hat{\nu}}\}$-stopping times, with

$$T_1^k \le T_2^k \le \cdots, \quad \forall k \in \mathbb{N}.$$

Put $T^k := \lim_{m\to\infty} T_m^k$, and observe that for every $k \in \mathbb{N}$ the limit

$$\hat{\mu}_t^k \equiv \lim_{m\to\infty} \hat{\nu}_{t\wedge T_m^k} \tag{C.3.74}$$

exists in $\mathcal{P}(\mathbb{R}^{d*})$. Also observe that for every $k \in \mathbb{N}$, $\omega \in \Omega$, $t \in [0,T)$ we have

$$\{\hat{\mu}_t^k(\omega) \in M_k\} \subset \{T^k(\omega) > t\}. \tag{C.3.75}$$

Fix some $t \in [0,T)$. Using the fact that for every $n \in \mathbb{N}$ the process $\{N_t^n\}$ is a corlol $\{\mathcal{F}_{t+}^{\hat{\nu}}\}$-martingale, we get

$$\mathrm{E}^{\hat{P}}[\hat{\nu}_T\,\phi_n] = \mathrm{E}^{\hat{P}}[\hat{\nu}_0\,\phi_n] + \mathrm{E}^{\hat{P}}\left[\int_0^T \hat{\nu}_s(\mathcal{L}^{\Delta}\phi_n)\,ds\right], \quad \forall k,\ m,\ n \in \mathbb{N}, \tag{C.3.76}$$

and, by the optional stopping theorem,

$$\mathrm{E}^{\hat{P}}[\hat{\nu}_{t\wedge T_m^k}\phi_n] = \mathrm{E}^{\hat{P}}[\hat{\nu}_0\phi_n] + \mathrm{E}^{\hat{P}}\left[\int_0^{t\wedge T_m^k} \hat{\nu}_s(\mathcal{L}^{\Delta}\phi_n)\,ds\right], \quad \forall k,\ m,\ n \in \mathbb{N}. \tag{C.3.77}$$

By letting $m \to \infty$ in (C.3.77), and using the dominated convergence theorem we get (recall (C.3.74))

$$\mathrm{E}^{\hat{P}}[\hat{\mu}_t^k\phi_n] = \mathrm{E}^{\hat{P}}[\hat{\nu}_0\phi_n] + \mathrm{E}^{\hat{P}}\left[\int_0^{t\wedge T^k} \hat{\nu}_s(\mathcal{L}^{\Delta}\phi_n)\,ds\right], \quad \forall t \in [0,T),\ k,\ n \in \mathbb{N}. \tag{C.3.78}$$

Since by Lemma D.3.146

$$\operatorname*{bp-lim}_{n\to\infty}(\phi_n, \mathcal{L}^{\Delta}\phi_n) = (I_{\mathbb{R}^d}, 0),$$

by letting $n \to \infty$ in (C.3.78) and using the dominated convergence theorem we get

$$\mathrm{E}^{\hat{P}}[\hat{\mu}_t^k(\mathbb{R}^d)] = \mathrm{E}^{\hat{P}}[\hat{\nu}_0(\mathbb{R}^d)], \quad \forall k \in \mathbb{N}. \tag{C.3.79}$$

Since $\hat{P}(\hat{\nu}_0(\mathbb{R}^d) = 1) = 1$, from (C.3.79) we get

$$\hat{P}(\hat{\mu}_t(\mathbb{R}^d) = 1) = 1, \tag{C.3.80}$$

while from (C.3.76)

$$\hat{P}(\hat{\nu}_T \in M_k) = 1, \tag{C.3.81}$$

To complete the proof, we exploit the structure of the stopping times T^k. Fixing some $k \in \mathbb{N}$, we observe by (C.3.75) and (C.3.80) that

$$\hat{P}(T^k > t) = 1,$$

which, by the definition of T^k, implies

$$\hat{P}(\hat{\nu}_s \in M_k,\ \forall s \in [0,t]) = 1.$$

Since $t \in [0,T)$ is arbitrary and

$$\{\hat{\nu}_s \in M_k,\ \forall s \in [0,T)\} = \bigcap_{i\in\mathbb{N}}\{\hat{\nu}_s \in M_k,\ \forall s \in [0, T - 1/i]\},$$

this implies

$$\hat{P}(\hat{\nu}_s \in M_k,\ \forall s \in [0,T)) = 1, \quad \forall k \in \mathbb{N},$$

which together with (C.3.81) implies

$$\hat{P}(\hat{\nu}_s \in M_k,\ \forall s \in [0,T]) = 1, \quad \forall k \in \mathbb{N}.$$

Together with (C.3.73) this in turn implies

$$\hat{P}(\hat{\nu}_s \in M, \ \forall s \in [0,T]) = 1,$$

as required.

□

Lemma C.3.114. *Let*

$$M_n := \{\mu \in \mathcal{P}(\mathbb{R}^{d*}) : \ \mu(\{\Delta\}) < 1/n\}, \quad n \in \mathbb{N}$$
$$M := \bigcap_{n\in\mathbb{N}} M_n = \{\nu \in \mathcal{P}(\mathbb{R}^{d*}) : \nu(\Delta) = 0\}.$$

Then the following holds:

(i) for every $n \in \mathbb{N}$, the set M_n is open in $\mathcal{P}(\mathbb{R}^{d*})$;

(ii) the set M is homeomorphic to $\mathcal{P}(\mathbb{R}^d)$.

Proof. To establish the first statement, we fix $n \in \mathbb{N}$ and show that the set

$$N := \{\mu \in \mathcal{P}(\mathbb{R}^{d*}) : \mu(\{\Delta\}) \geq 1/n\}$$

is closed in $\mathcal{P}(\mathbb{R}^{d*})$. To this end, fix a sequence $(\mu_n)_{n=1}^{\infty} \subset N$ converging weakly to some $\mu \in \mathcal{P}(\mathbb{R}^{d*})$, and observe that, since $\{\Delta\}$ is closed in $\mathbb{R}^{d*}$, by the Portmanteau theorem (e.g. Theorem 3.3.1 of Ethier and Kurtz [8]), we have

$$\limsup_{n\to\infty} \mu_n(\{\Delta\}) \leq \mu(\{\Delta\}).$$

Therefore, $\mu(\{\Delta\}) \geq 1/n$, which implies $\mu \in N$, hence the closedness of N.

The second claim is immediate consequence of the definition of weak convergence (see Remark D.1.117) and the definition of M. □

Appendix D

Background Material

D.1 Convergence of Measures in Metric Spaces

Throughout this section $(E, \mathcal{U})$ is a Polish space, that is, a topological space homeomorphic to a complete separable metric space.

Definition D.1.115. The *weak topology* $\mathcal{T}$ on $\mathcal{M}^+(E)$ is generated by the subbase

$$\left\{\mu \in \mathcal{M}^+(E) : \left|\int_E f\,d\mu - \int_E f\,d\mu_0\right| < \epsilon\right\}, \quad \epsilon \in (0,\infty),\ f \in \bar{C}(E),\ \mu_0 \in \mathcal{M}^+(E).$$

Remark D.1.116. Since E is a Polish space, it can be established that $(\mathcal{M}^+(E), \mathcal{T})$ is also a Polish space (see Prohorov [29]). We shall usually write $\mathcal{M}^+(E)$ in place of $(\mathcal{M}^+(E), \mathcal{T})$. It is also straightforward to check that $\mathcal{P}(E)$ is a closed set in $\mathcal{M}^+(E)$.

Remark D.1.117. From Definition D.1.115 it follows that the sequence $(\mu_n)_{n=1}^{\infty} \subset \mathcal{M}^+(E)$ converges weakly to $\mu \in \mathcal{M}^+(E)$ iff

$$\lim_{n\to\infty} \mu_n f = \mu f$$

for every $f \in \bar{C}(E)$.

Remark D.1.118. Using the fact that E is Polish, in view of Remark D.1.116 we see that in order to show that a set $\Pi \subset \mathcal{P}(E)$ is relatively compact, it is enough to show that Π is *sequentially* relatively compact, namely that each sequence $(\mu_n)_{n=1}^{\infty} \subset \Pi$ contains a convergent subsequence (in $\mathcal{P}(E)$).

Definition D.1.119. Let $(X_\lambda)_{\lambda\in\Lambda}$ be a collection of mappings from (possibly different) probability spaces $(\Omega_\lambda, \mathcal{F}_\lambda, P_\lambda)$ into E. We say that the collection $(X_\lambda)_{\lambda\in\Lambda}$ is relatively compact if the collection of probability measures $(\mathcal{L}(X_\lambda)_{\lambda\in\Lambda} \subset \mathcal{P}(E)$ is relatively compact.

Remark D.1.120. For a sequence of mappings $(X_n)_{n=0}^{\infty}$ defined on (possibly different) probability spaces $(\Omega_n, \mathcal{F}_n, P_n)$ taking values in E we say that $(X_n)_{n=0}^{\infty}$ *converges weakly* to X_0 if

$$\lim_{n\to\infty} \mathcal{L}_{P_n}(X_n) = \mathcal{L}_{P_0}(X_0) \quad \text{in } \mathcal{P}(E)$$

(in the literature this is frequently denoted by $X_n \Rightarrow X_0$). When there is no possibility of confusion, we shorten the notation to

$$\lim_{n\to\infty} \mathcal{L}(X_n) = \mathcal{L}(X_0) \quad \text{in } \mathcal{P}(E).$$

Definition D.1.121. The set $M \subset \bar{C}(E)$ *separates points* in E when the equality

$$f(x) = f(y), \quad \forall\, f \in M,$$

for some $x, y \in E$, implies $x = y$.

Definition D.1.122. The set $M \subset \bar{C}(E)$ is *separating* when the equality

$$Pf = Qf, \quad \forall f \in M,$$

for some $P, Q \in \mathcal{P}(E)$, implies $P = Q$.

Definition D.1.123. The set $M \subset \bar{C}(E)$ is *separating for* $\mathcal{M}^+(E)$ when the equality

$$\mu f = \nu f, \quad \forall f \in M,$$

for some $\mu, \nu \in \mathcal{M}^+(E)$, implies $\mu = \nu$.

Remark D.1.124. We see that, if $M \subset \bar{C}(E)$ is separating for $\mathcal{M}^+(E)$, then of course M is separating; and, if M is separating, then M separates points in E.

Definition D.1.125. The set $M \subset \bar{C}(E)$ is said to *strongly separate points* in E when the convergence

$$\lim_n f(x_n) = f(x_0), \quad \forall\, f \in M,$$

for some sequence $(x_n)_{n=0}^\infty \subset E$, implies $\lim_n x_n = x_0$.

Definition D.1.126. A set $M \subset \bar{C}(E)$ is called *convergence determining* when the convergence

$$\lim_{n\to\infty} P_n f = P_0 f, \quad \forall f \in M,$$

for some sequence $(P_n)_{n=0}^\infty \subset \mathcal{P}(E)$, implies weak convergence of $(P_n)_{n=1}^\infty$ to P_0.

Definition D.1.127. A set $M \subset \bar{C}(E)$ is called *convergence determining for* $\mathcal{M}^+(E)$ when the convergence

$$\lim_{n\to\infty} \mu_n f = \mu_0 f, \quad \forall f \in M,$$

for some sequence $(\mu_n)_{n=0}^\infty \subset \mathcal{M}^+(E)$, implies weak convergence of $(\mu_n)_{n=1}^\infty$ to μ_0.

Remark D.1.128. We see that, if $M \subset \bar{C}(E)$ is convergence determining for $\mathcal{M}^+(E)$, then of course M is convergence determining. Moreover, if M is convergence determining, it also strongly separates points in E.

Fact D.1.129. *The set of functions $C_c^\infty(\mathbb{R}^d)$ is (i) convergence determining, and (ii) separating for $\mathcal{M}^+(\mathbb{R}^d)$.*

Proof. (i) It follows from Fact D.3.144 and Lemma D.3.143 that $C_c^\infty(\mathbb{R}^d)$ is convergence determining. (ii) follows from Problem 5.4.25 on page 325 of Karatzas and Shreve [16]. $\square$

Definition D.1.130. We say that a family of probability measures $\Pi \subset \mathcal{P}(E)$ is *relatively compact* if each sequence $(P_n)_{n=1}^{\infty} \subset \Pi$ contains a weakly convergent subsequence.

Next, we present a pair of theorems that, when coupled together, provide a very powerful method for establishing relative compactness of a sequence of continuous processes taking values in a general complete and separable metric space. These results are simplified versions of, respectively, Theorem 3.9.1 and Theorem 3.9.4 of Ethier and Kurtz [8]. Under slightly different conditions the second result was also given in Jakubowski [14].

Remark D.1.131. Ethier and Kurtz [8] give both results in the different setting in which all processes under consideration are *corlol*, rather than continuous. Since $C_E[0,\infty)$ is a closed subset of $D_E[0,\infty)$ (see e.g. Problem 3.11.25(c) of Ethier and Kurtz [8]), the continuous case we treat follows easily from their more general results by the Portmanteau theorem (e.g. Theorem 3.1.1 of Ethier and Kurtz [8]).

Theorem D.1.132. *Let (E,r) be complete and separable, and let $\{X_t^n,\ t \in [0,\infty)\}$ be a processes with sample paths in $C_E[0,\infty)$ defined on some probability space $(\Omega^n, \mathcal{F}^n, P^n)$, for $n = 1,2,\ldots$. Suppose that the following* compact containment condition *holds: for every $\eta > 0$ and $q \in (0,\infty)$ there there exists a compact set $K_{\eta,q} \subset E$ such that*

$$\inf_{n\in\mathbb{N}} P^n(\{X_t^n \in K_\eta,\ \forall t \in [0,q]\}) \geq 1-\eta. \tag{D.1.1}$$

Let U be a dense subset of $\bar{C}(E)$ in the topology of uniform convergence on compact sets. Then, if $\{\mathcal{L}(f(X^n)),\ n \in \mathbb{N}\}$ is relatively compact in $\mathcal{P}(C_{\mathbb{R}}[0,\infty))$ for every $f \in U$, the sequence $\{\mathcal{L}(X^n),\ n \in \mathbb{N}\}$ is relatively compact in $\mathcal{P}(C_E[0,\infty))$.

Theorem D.1.133. *Let E be a metric space, and let $\{X_t^n,\ t \in [0,\infty)\}$, $n \in \mathbb{N}$ be a sequence of processes with sample paths in $C_E[0,\infty)$ such that each $\{X_t^n\}$ is defined* *on a filtered probability space $(\Omega^n, \mathcal{F}^n, \{\mathcal{F}_t^n\}, P^n)$ and adapted to the filtration $\{\mathcal{F}_t^n\}$. Let C_a be a subalgebra of $\bar{C}(E)$ such that for every $q > 0$ there exists a sequence of $\mathbb{R}^2$-valued $\{\mathcal{F}_t^n\}$–progressively measurable processes $\{(U_t^n, V_t^n),\ t \in [0,\infty)\}$, $n \in \mathbb{N}$, with*

1. $\sup_{t\in[0,\infty)} \mathrm{E}^{P^n}[|U_t^n| + |V_t^n|] < \infty$;
2. *for every $n \in \mathbb{N}$ the process $U_t^n - \int_0^t V_s^n\, ds,\ t \in [0,\infty)$ is an $\{\mathcal{F}_t^n\}$–martingale;*
3. *for every $\phi \in C_a$*

$$\limsup_{n\to\infty} \mathrm{E}^{P^n}\left[\sup_{t\in[0,q]} |U_t^n - \phi(X_t^n)|\right] = 0, \tag{D.1.2}$$

and

$$\limsup_{n\to\infty} \mathrm{E}^{P^n}\left[\operatorname*{ess\,sup}_{t\in[0,q]} |V_t^n|\right] < \infty. \tag{D.1.3}$$

Then the sequence $\{\mathcal{L}(\phi(X^n)),\ n \in \mathbb{N}\}$ is relatively compact in $\mathcal{P}(C_{\mathbb{R}}[0,\infty))$ for every $\phi \in C_a$.

D.2 Martingale Problem

The Martingale Problem was originally formulated in the context of finite-dimensional diffusion processes by Stroock and Varadhan [35]. Here, following Ethier and Kurtz [8], we present a more general formulation which is essential for nonlinear filtering.

Definition D.2.134. Suppose that (E, r) is a metric space, $\mu \in \mathcal{P}(E)$ is a probability measure, and that $(\mathcal{A}, \mathcal{D}(\mathcal{A}))$ is some (not necessarily linear) operator from some domain $\mathcal{D}(A) \subset B(E)$ into $B(E)$. Then an E-valued process $\{X_t,\ t \in [0,\infty)\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ is a *solution of the martingale problem for* $(\mathcal{A}, \mu)$ if

1. $\mathcal{L}(X(0)) = \mu$;

2. there is a filtration $\{\mathcal{F}_t\}$ in $\mathcal{F}$ such that $\{(X_t, \mathcal{F}_t)\}$ is progressively measurable;

3. for each $f \in \mathcal{D}(A)$ the process

$$f(X_t) - \int_0^t \mathcal{A}f(X_s)\, ds, \quad \forall t \in [0, \infty) \tag{D.2.4}$$

is an $\{\mathcal{F}_t\}$-martingale.

If the initial measure is not important, we simply say that $\{X_t\}$ *solves the martingale problem for* $\mathcal{A}$.

Remark D.2.135. In many instances the solution $\{X_t\}$ of the martingale problem has corlol (or even continuous) paths, in which case the filtration $\{\mathcal{F}_t\}$ in Definition D.2.134 can be taken as $\mathcal{F}_t := \sigma\{X_s,\ 0 \le s \le t\}$.

Remark D.2.136. Suppose that (E, r), μ, and $(\mathcal{A}, \mathcal{D}(\mathcal{A}))$ are as in Definition D.2.134. Then a probability measure $P \in \mathcal{P}(D_E[0,\infty))$ $[P \in \mathcal{P}(C_E[0,\infty))]$ is called a solution of the martingale problem for $(\mathcal{A}, \mu)$ when the conditions of Definition D.2.134 hold, with $\{X_t\}$ denoting the canonical process on $D_E[0,\infty)$ $[C_E[0,\infty)]$ and

$$\mathcal{F}_t := \sigma\{X_s,\ 0 \le s \le t\}.$$

Definition D.2.137. The martingale problem for $\mathcal{A}$ is said to be *well posed* if, for each $\mu \in \mathcal{P}(E)$, the following hold:

1. There exists a solution of the martingale problem for $(\mathcal{A}, \mu)$;

2. Any two solutions of the martingale problem for $(\mathcal{A}, \mu)$ have identical finite-dimensional distributions.

Theorem D.2.138 (Theorem 2.1 of Bhatt and Karandikar [2]). *Suppose that E is a complete separable metric space, and $\mathcal{A} \subset \bar{C}(E) \times \bar{B}(E)$ is a linear operator with domain $\mathcal{D}(\mathcal{A})$ having the following properties:*

1. *the corlol martingale problem for $(\mathcal{A}, \delta_x)$ is well-posed for each $x \in E$;*
2. *there exists a countable set $\mathfrak{B} \subset \mathcal{D}(\mathcal{A})$ such that $\{(f, \mathcal{A}f) : f \in \mathcal{D}(\mathcal{A})\}$ is a subset of the bp-closure of $\{(f, \mathcal{A}f) : f \in \mathfrak{B}\}$.*

Then the corlol martingale problem for $\mathcal{A}$ is well-posed.

Definition D.2.139. Suppose that E is a metric space. An operator $(\mathcal{A}, \mathcal{D}(\mathcal{A}))$ on $B(E)$ (that is, $\mathcal{A} \subset B(E) \times B(E)$) is called *conservative* when $(1,0)$ is contained in the bp-closure of (the graph of) $\mathcal{A}$.

Theorem D.2.140 (Theorem 4.3.8 of Ethier and Kurtz [8]). *Let $(\hat{E}, r)$ be a metric space and let $A \subset B(\hat{E}) \times B(\hat{E})$. Let $E \subset \hat{E}$ be open, and suppose that $\{X_t\}$ is a solution of the martingale problem for A with corlol sample paths. Suppose $(\chi_E, 0)$ is in the bp-closure of $A \cap (\bar{C}(\hat{E}) \times B(\hat{E}))$. If $P(\{X_0 \in E\}) = 1$, then $P(\{X \in D_E[0,\infty)\}) = 1$. i.e. $\{X_t\}$ is E-valued with corlol sample paths a.s.*

Theorem D.2.141 (Theorem 4.8.10 of Ethier and Kurtz [8]). *Let (E, r) be complete and separable metric space. Let $A \subset \bar{C}(E) \times \bar{C}(E)$ and $\nu \in \mathcal{P}(E)$, and suppose the following:*

1. *uniqueness holds for the $C_E[0,\infty)$ martingale problem for (A, ν);*
2. *for each $n \in \mathbb{N}$ there is a continuous E-valued adapted process $\{X_t^n,\ t \in [0,\infty)\}$ on a filtered probability space $(\Omega^n, \mathcal{F}^n, \{\mathcal{F}_t^n\}, P^n)$;*
3. *the sequence $\{\mathcal{L}(X^n),\ n \in \mathbb{N}\}$ is relatively compact in $\mathcal{P}(C_E[0,\infty))$;*
4. *the sequence $\mathcal{L}(X_0^n)$ converges to ν in $\mathcal{P}(E)$;*
5. *for each $(f, g) \in A$, $T \in [0,\infty)$, and $n \in \mathbb{N}$, there exists a pair $\{(U_t^n, V_t^n)\}$ of real-valued processes defined on $(\Omega^n, \mathcal{F}^n, P^n)$ that are progressively measurable with respect to $\{\mathcal{F}_t^n\}$, such that*

$$U_n(t) - \int_0^t V_n(s)\, ds \quad \textit{is an } \{\mathcal{F}_t^n\}\textit{-martingale} \quad \forall n \in \mathbb{N},$$

and

$$\sup_n \sup_{0\le s\le T} \mathrm{E}[|U_s^n|] < \infty,$$
$$\sup_n \sup_{0\le s\le T} \mathrm{E}[|V_s^n|] < \infty,$$
$$\lim_{n\to\infty} \mathrm{E}\left[(U_t^n - f(X_t^n))\prod_{i=1}^k h_i(X_{t_i}^n)\right] = 0,$$
$$\lim_{n\to\infty} \mathrm{E}\left[(V_t^n - g(X_t^n))\prod_{i=1}^k h_i(X_{t_i}^n)\right] = 0,$$

for finitely many $0 \le t_1 < t_2 < \ldots < t_k \le t \le T$ *and* $h_1, \ldots, h_k \in \bar{C}(E)$.

Then there exists a solution $\{X_t\}$ *of the* $C_E[0,\infty)$ *martingale problem for* (A,ν)*, and* $\mathcal{L}(X^n)$ *converges to* $\mathcal{L}(X)$ *in* $\mathcal{P}(C_E[0,\infty))$.

D.3 Miscellaneous Technical Results

Lemma D.3.142 (Problem 2.9.22 of Ethier and Kurtz [8]). *Let* $\{X_t,\ t\in[0,\infty)\}$ *and* $\{Y_t,\ t\in[0,\infty)\}$ *be processes on a filtered probability space* $(\Omega,\mathcal{F},\{\mathcal{F}_t\},P)$ *such that* $\{X_t\}$ *is* $\{\mathcal{F}_t\}$*-martingale and* $\{Y_t\}$ *is* $\{\mathcal{F}_t\}$*-adapted. Then, if*

$$\mathrm{E}\left[\int_0^t |X_t Y_s|\,ds\right] < \infty, \quad \forall t\in[0,\infty),$$

the process

$$Z_t := X_t\int_0^t Y_s\,ds - \int_0^t X_s Y_s\,ds, \quad t\in[0,\infty)$$

is an $\{\mathcal{F}_t\}$*-martingale.*

Proof. First note that, since $\{|X_t|,\ t\in[0,\infty)\}$ is a submartingale, we have

$$\int_0^t \mathrm{E}[|X_sY_s|]\,ds \le \int_0^t \mathrm{E}[\mathrm{E}[|X_t||\mathcal{F}_s]|Y_s|]\,ds \le \int_0^t \mathrm{E}[|X_tY_s|]\,ds < \infty.$$

Therefore, by the Fubini's theorem for conditional expectations (e.g. Proposition 2.4.6 of Ethier and Kurtz [8]) we have a.s. for every $t > s$

$$\begin{aligned}
\mathrm{E}[Z_t|\mathcal{F}_s] &= \mathrm{E}\left[X_t \int_0^t Y_u \, du \middle| \mathcal{F}_s \right] - \mathrm{E}\left[\int_0^t X_u Y_u \, du \middle| \mathcal{F}_s \right] \\
&= \int_0^s \mathrm{E}[X_t Y_u|\mathcal{F}_s] \, du + \int_s^t \mathrm{E}[X_t Y_u|\mathcal{F}_s] \, du \\
&- \int_0^s \mathrm{E}[X_u Y_u|\mathcal{F}_s] \, du - \int_s^t \mathrm{E}[X_u Y_u|\mathcal{F}_s] \, du \\
&= \int_0^s X_s Y_u \, du + \int_s^t \mathrm{E}[X_t Y_u|\mathcal{F}_s] \, du - \int_0^s X_u Y_u \, du - \int_s^t \mathrm{E}[X_u Y_u|\mathcal{F}_s] \, du.
\end{aligned} \tag{D.3.5}$$

But, since $\{X_t\}$ is an $\{\mathcal{F}_t\}$-martingale

$$\begin{aligned}
\int_s^t \mathrm{E}[X_t Y_u|\mathcal{F}_s] \, du = \int_s^t \mathrm{E}[\mathrm{E}[X_t Y_u|\mathcal{F}_u]|\mathcal{F}_s] \, du &= \int_s^t \mathrm{E}[Y_u \, \mathrm{E}[X_t|\mathcal{F}_u]|\mathcal{F}_s] \, du \\
&= \int_s^t \mathrm{E}[Y_u X_u|\mathcal{F}_s] \, du,
\end{aligned}$$

whence from (D.3.5)

$$\mathrm{E}[Z_t|\mathcal{F}_s] = Z_s \quad \text{a.s..}$$

□

Lemma D.3.143 (Problem 3.11.11 of Ethier and Kurtz [8]**).** *Let E be locally compact separable metric space. Then $M \subset \hat{C}(E)$ is convergence determining iff M is dense in $\hat{C}(E)$ in the supremum norm.*

Proof. Since $M \subset \bar{C}(E)$, the necessity is trivial. For sufficiency, fix some $(P_n)_{n=0}^{\infty} \subset \mathcal{P}(E)$ such that

$$\lim_{n\to\infty} P_n f = P_0 f, \quad \forall f \in M, \tag{D.3.6}$$

and let $\epsilon > 0$, $f \in \hat{C}(E)$ be arbitrary. Since M is dense in $\hat{C}(E)$, there exists $g \in M$ such that

$$\|f - g\| < \epsilon/2. \tag{D.3.7}$$

Since

$$\limsup_{n\to\infty} \left| \int f\, dP_n - \int f\, dP_0 \right| \\ \leq \limsup_{n\to\infty} \left(\int \|f-g\|\, dP_n + \int \|f-g\|\, dP_0 \right) + \lim_{n\to\infty} \left| \int g\, dP_n - \int g\, dP_0 \right|,$$

in view of (D.3.6) and (D.3.7) we get

$$\limsup_{n\to\infty} \left| \int f\, dP_n - \int f\, dP_0 \right| < \epsilon.$$

Therefore, since $\epsilon > 0$ is arbitrary, we have

$$\lim_{n\to\infty} \int f\, dP_n = \int f\, dP_0, \quad \forall f \in \hat{C}(E),$$

whence the result follows as $\hat{C}(E)$ is convergence determining (see Proposition 3.4.4 of Ethier and Kurtz [8]). □

Fact D.3.144. *The set $C_c^\infty(\mathbb{R}^d)$ is dense in $\hat{C}(\mathbb{R}^d)$ in the supremum norm.*

Proof. Since $C_c^\infty(\mathbb{R}^d)$ is an algebra which separates points and vanishes nowhere, the result follows by the Stone-Weierstrass theorem (see Theorem 4.52 of Folland [9]). □

Lemma D.3.145 (Problem 2.9.29 of Ethier and Kurtz [8]). *Let E be a metric space and let*

$\{X_t,\ t \in [0,\infty)\}$ be a right-continuous E-valued process on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$, adapted to $\{\mathcal{F}_t\}$. Let f, g, $h \in B(E)$ be such that

$$M_f(t) := f(X(t)) - f(X(0)) - \int_0^t g(X(s))\, ds, \quad t \in [0,\infty)$$

and

$$f^2(X(t)) - f^2(X(0)) - \int_0^t h(X(s))\,ds, \quad t \in [0,\infty) \tag{D.3.8}$$

are $\{\mathcal{F}_t\}$-martingales. Then

$$M_f^2(t) - \int_0^t \big(h(X(s)) - 2f(X(s))g(X(s))\big)\,ds, \quad t \in [0,\infty) \quad \textit{is an } \{\mathcal{F}_t\}\textit{-martingale.} \tag{D.3.9}$$

Proof. Let us for two processes $\{U_t\}$ and $\{V_t\}$ write $U_t \cong V_t$ if $U_t - V_t$ is an $\{\mathcal{F}_t\}$-martingale. First note that it is enough to show that (D.3.9) holds with the martingale $\hat{M}_f(t) := M_f(t) - f(X(0))$, $t \in [0,\infty)$ in place of M_f, as

$$\hat{M}_f^2(t) = M_f^2(t) - 2f(X(0))M_f(t) + f^2(X(0)) \cong M_f^2(t). \tag{D.3.10}$$

By Lemma D.3.142, we have

$$\begin{aligned}
\hat{M}_f^2(t) = &\left[f(X(t)) - \int_0^t g(X(s))\,ds\right]\hat{M}_f(t) \cong f(X(t))\hat{M}_f(t) - \int_0^t \hat{M}_f(s)g(X(s))\,ds \\
=&f^2(X(t)) - f(X(t))\int_0^t g(X(s))\,ds - \int_0^t \hat{M}_f(s)g(X(s))\,ds \\
=&f^2(X(t)) - \hat{M}_f(t)\int_0^t g(X(s))\,ds - \left(\int_0^t g(X(s))\,ds\right)^2 - \int_0^t \hat{M}_f(s)g(X(s))\,ds \\
\cong&f^2(X(t)) - 2\int_0^t \hat{M}_f(s)g(s)\,ds - \left(\int_0^t g(X(s))\,ds\right)^2.
\end{aligned} \tag{D.3.11}$$

Using the fact that the process in (D.3.8) is an $\{\mathcal{F}_t\}$-martingale, from (D.3.11) we get

$$\begin{aligned}
\hat{M}_f^2(t) \cong &\int_0^t \big(h(X(s)) - 2f(X(s))g(X(s))\big)ds + \\
&2\int_0^t g(X(s))\int_0^s g(X(u))\,du\,ds - \left(\int_0^t g(X(s))\,ds\right)^2.
\end{aligned} \tag{D.3.12}$$

But integration by parts gives

$$\begin{aligned}\left(\int_0^t g(X(s))\,ds\right)^2 &= \int_0^t g(X(s))\int_0^s g(X(u))\,du\,ds + \int_0^t g(X(s))\int_0^s g(X(u))\,du\,ds\\ &= 2\int_0^t g(X(s))\int_0^s g(X(u))\,du\,ds,\end{aligned}$$

so from (D.3.12)

$$\hat{M}_f(t)^2 \cong \int_0^t \big(h(X(s)) - 2f(X(s))g(X(s))\big)\,ds.$$

□

Lemma D.3.146 (Problem 4.11.12 of Ethier and Kurtz [8]). *Consider the diffusion operator*

$$\mathcal{A} := \sum_{i=1}^d b^i(x)\partial_i + \frac{1}{2}\sum_{i,j=1}^d a^{ij}(x)\partial_i\partial_j,\ \ \mathcal{D}(\mathcal{A}) := C_c^\infty(\mathbb{R}^d),$$

where a^{ij}, $1 \le i,j \le d$ and b^j, $1 \le j \le d$ are Borel measurable functions satisfying

$$|a^{ij}(x)| \le C(1+|x|^2),\ \ |b^i(x)| \le C(1+|x|),\quad \forall x \in \mathbb{R}^d,\ 1 \le i,j \le d \tag{D.3.13}$$

for some constant $C \in [0,\infty)$. Let $g \in C_c^\infty(\mathbb{R}^d)$ be such that $\|g\| = 1$, $g(x) = 1$ for $|x| \le 1$, and $g(x) = 0$ for $|x| \ge 2$, and put

$$g_n(x) := g(x/n),\quad \forall x \in \mathbb{R}^d,\ n \in \mathbb{N}.$$

Then

$$\operatorname*{bp-lim}_{n\to\infty}\big(g_n, \mathcal{A}g_n\big) = (1,0),$$

i.e. $\mathcal{A}$ is conservative.

Proof. We have

$$\mathcal{A}g_k(x) = \sum_{i=1}^d b^i(x)\partial_i g_k(x) + \frac{1}{2}\sum_{i,j=1}^d a^{ij}(x)\partial_i\partial_j g_k(x),\quad x \in \mathbb{R}^d,\ k \in \mathbb{N},$$

and so, since $\mathcal{A}g_k(x) = 0$, $\forall |x| \le k$, it follows that

$$\lim_{k\to\infty} \big(g_k(x), \mathcal{A}g_k(x)\big) = (1,0), \quad \forall x \in \mathbb{R}^d. \tag{D.3.14}$$

Furthermore, by (D.3.13) for every $k \in \mathbb{N}$

$$|\mathcal{A}g_k(x)| \le \sum_{i=1}^{d} \frac{C(1+|x|)}{k} |\partial_i g(x/k)| + \frac{1}{2} \sum_{i,j=1}^{d} \frac{C(1+|x|^2)}{k^2} |\partial_i \partial_j g(x/k)|, \quad \forall x \in \mathbb{R}^d,$$

so, from the fact that

$$\partial_i g(x/k) = 0, \quad \partial_i \partial_j g(x/k) = 0, \quad \forall |x| \ge 2k,$$

we get

$$\begin{aligned} \|\mathcal{A}g_k\| &\le \sum_{i=1}^{d} \frac{C(1+2k)}{k} \|\partial_i g\| + \frac{1}{2} \sum_{i,j=1}^{d} \frac{C(1+4k^2)}{k^2} \|\partial_i \partial_j g\| \\ &\le 3C \sum_{i=1}^{d} \|\partial_i g\| + 3C \sum_{i,j=1}^{d} \|\partial_i \partial_j g\| < \infty, \quad \forall k \in \mathbb{N}. \end{aligned} \tag{D.3.15}$$

Thus, the result follows from (D.3.14), (D.3.15), and the fact that $\sup_k \|g_k\| < \infty$. $\square$

Lemma D.3.147. *Suppose ν is a finite signed measure on S, and for some $g \in C^{1,0}(\mathbb{R}^d \times S)$ let*

$$\phi(x) := \int_S g(x, z')\, \nu(dz'), \quad x \in \mathbb{R}^d.$$

Then $\phi \in C^1(\mathbb{R}^d)$, and

$$\partial_j \phi(x) = \int_S \partial_j g(x, z')\, \nu(dz'), \quad 1 \le j \le d,\ x \in \mathbb{R}^d.$$

Proof. This is essentially Theorem 2.27 of Folland [9]: one has to split ν using the Hahn decomposition, and then apply the theorem separately to the integrals corresponding to the positive and negative parts of ν. $\square$

The next result is easily established using the Dynkin class theorem:

Lemma D.3.148 (Problem 1.5.7 of Stroock and Varadhan [35]). *Suppose that $(\Omega, \mathcal{F}, P)$ is a probability space, $\mathcal{G} \subset \mathcal{F}$ is a conditioning sub-σ-algebra, $(M, \mathcal{M})$ is a measurable space, and $\Psi : \Omega \times M \to \mathbb{R}$ is $\mathcal{F} \times \mathcal{M}$-measurable, such that*

$$\sup_{y \in M} \mathrm{E}\,|\Psi(\cdot, y)| < \infty.$$

Then there exists a $\mathcal{G} \times \mathcal{M}$-measurable mapping $\Phi : \Omega \times M \to \mathbb{R}$ such that, for each $y \in M$, we have

$$\Phi(\cdot, y) = \mathrm{E}[\Psi(\cdot, y)|\mathcal{G}] \quad a.s.$$

In addition, suppose that $\phi : \Omega \to M$ is $\mathcal{G}/\mathcal{M}$-measurable, and put

$$X(\omega) := \Psi(\omega, \phi(\omega)), \qquad Y(\omega) := \Phi(\omega, \phi(\omega)), \quad \forall\, \omega \in \Omega.$$

If $\mathrm{E}\,|X| < \infty$, then

$$\mathrm{E}[X|\mathcal{G}] = Y, \quad a.s.$$

Appendix E

Glossary of Notation and Terminology

Let (E, r) be a separable metric space with the corresponding Borel σ-algebra $\mathcal{B}(E)$, let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $T \in (0, \infty)$ be fixed. We adopt the following notation:

I Designated Collections of Real-Valued Functions

$B(E)$	Banach space of all real-valued uniformly bounded, $\mathcal{B}(E)$-measurable functions, with the supremum norm: $\|\|f\|\| = \sup_{x\in E} \|f(x)\|$, $\forall f \in B(E)$.
$\bar{C}(E)$	Banach space of all bounded continuous real-valued functions on E with the supremum norm.
$\hat{C}(E)$	Banach space of all continuous real-valued functions vanishing at infinity, with the supremum norm (defined for locally compact E).
$C_c(E)$	set of all continuous real-valued functions on E with compact support.

$C^n(\mathbb{R}^d)$	set of continuous real-valued functions on $\mathbb{R}^d$ having continuous derivatives of every order up to some integer $n \geq 1$.
$C_c^\infty(\mathbb{R}^d)$	collection of all infinitely differentiable functions in $C_c(\mathbb{R}^d)$.
$C_c^{n,0}(\mathbb{R}^d \times E)$	space of all continuous real-valued functions on $\mathbb{R}^d \times E$ with compact support and whose all n-th order partial derivatives with respect to the first co-ordinate exist and are continuous.

II Designated Collections of Path Spaces

$C_E[0,T]$	space of continuous functions $f : [0,T] \to E$, with the topology of uniform convergence.
$C_E[0,\infty)$	space of continuous functions $f : [0,\infty) \to E$, with the topology of uniform convergence on compacts.
$D_E[0,T]$	space of functions $f : [0,T] \to E$ that are right-continuous and have left limits equipped with the Skorohod topology.
$D_E[0,\infty)$	space of functions $f : [0,\infty) \to E$ that are right-continuous and have left limits equipped with the Skorohod topology.

III Notation For Measures and Integrals

$\mathcal{M}^+(E)$	the topological space of all all *finite positive* measures on $\mathcal{B}(E)$ with the topology of weak convergence.
$\mathcal{P}(E)$	the subspace of $\mathcal{M}^+(E)$ consisting of all *probability* measures on $\mathcal{B}(E)$

with the topology of weak convergence.

$\mathcal{L}(X)$ the law of the mapping $X : \Omega \to E$, that is,

$\mathcal{L}(X)(\Gamma) := P(X \in \Gamma),\ \Gamma \in \mathcal{B}(E)$.

μf the integral of $f \in B(E)$ with respect to $\mu \in \mathcal{P}(E)$, that is,

$\mu f := \int_E f(x)\,\mu(dx)$.

IV Filtrations Generated by a Process

Let $\{X_t,\ t \in [0,T]\}$ be a given E-valued stochastic process on some complete probability space $(\Omega, \mathcal{F}, P)$. We denote the collection of all P-null events in $\mathcal{F}$ by $\mathcal{N}(P)$, namely

$$\mathcal{N}(P) := \{N \in \mathcal{F} :\ P(N) = 0\},$$

and define

$$\begin{aligned}\mathcal{F}_t^X &:= \sigma\{X_u :\ 0 \le u \le t\} \vee \mathcal{N}(P),\ \ \forall t \in [0,T];\\ \mathcal{F}_{t+}^X &:= \textstyle\bigcap_{\epsilon>0} \mathcal{F}_{t+\epsilon},\ \ \forall t \in [0,T).\end{aligned}$$

V Miscellaneous Notation and Terminology

In the context of Euclidean spaces, $|\cdot|$ will denote the Euclidean norm, $|x| = \sqrt{\sum_{i=1}^d x_i^2}$, $\forall x \in \mathbb{R}^d$.

The one-point compactification of $\mathbb{R}^d$ will be denoted by $\mathbb{R}^{d*}$.

For the positive integers q, r the symbol $\mathbb{R}^{q\times q}$ will denote the set of all $q \times r$ matrices with real entries. Likewise, $\mathbb{S}_+^{q\times q}$ will denote the collection of all members of $\mathbb{R}^{q\times q}$

that are symmetric positive semidefinite, and $\mathbb{S}_{++}^{q\times q}$ will denote all members of $\mathbb{S}_{+}^{q\times q}$ that are strictly positive definite.

A process $\{Z_t,\ t \in [0,T]\}$, defined on a probability space $(\Omega, \mathcal{F}, P)$ taking values in a metric space E is said to be *corlol* (continuous-on-right-limits-on-left) if

$$Z_{\cdot}(\omega) \in D_E[0,T],\ \forall \omega \in \Omega.$$

If V is a vector space over $\mathbb{R}$ and $U \subset V$ is arbitrary, span(U) will denote the linear span of elements in U.

The symbol $\nabla_x f$ will denote the row vector of x-partial derivatives of the function f; if f is only the function of x, the subscript "x" will be omitted.

The symbol Δ will denote the Laplacian operator.

Bibliography

[1] A. G. Bhatt, G. Kallianpur, and R. L. Karandikar. Uniqueness and robustness of solution of measure-valued equations of nonlinear filtering. *Ann. Probab.*, 23:1895–1938, 1995.

[2] A. G. Bhatt and R. L. Karandikar. Invariant measures and evolution equations for Markov processes characterized via martingale problems. *Ann. Probab.*, 21:2246–2268, 1993.

[3] A. G. Bhatt and R. L. Karandikar. Weak convergence to a Markov process: the martingale approach. *Probab. Theory Relat. Fields*, 96:335–351, 1993.

[4] A. G. Bhatt and R. L. Karandikar. Evolution equations for Markov processes: application to the white-noise theory of filtering. *Appl. Math. Optim.*, 31:327–348, 1995.

[5] G. Blankenship and G. C. Papanicolaou. Stability and control of stochastic systems with wide-band noise disturbances I. *SIAM J. Appl. Math.*, 34:437–476, 1978.

[6] D. A. Dawson. Measure-Valued Markov Processes. In *Lecture Notes in Mathematics*, volume 1541, pages 1–260. Springer-Verlag, 1993.

[7] J. Dugundji. *Topology*. Allyn and Bacon, 1966.

[8] S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence.* J. Wiley & Sons, 1986.

[9] G. B. Folland. *Real Analysis: Modern Techniques and Their Applications.* J. Wiley & Sons, 1984.

[10] M. Fujisaki, G. Kallianpur, and H. Kunita. Stochastic differential equations for non-linear filtering problem. *Osaka J. Math.*, 9:19–40, 1972.

[11] E. M. Goggin. Convergence in distribution of conditional expectations. *Ann. Probab.*, 22:1097–1114, 1994.

[12] O. Hijab. Partially observed control of Markov processes, I. *Stochastics*, 28:123–144, 1989.

[13] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes.* North-Holland, 1989.

[14] A. Jakubowski. On the Skorokhod topology. *Ann. Inst. Henri Poincaré*, 22:263–285, 1986.

[15] G. Kallianpur. *Stochastic Filtering Theory.* Springer–Verlag, 1980.

[16] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus.* Springer–Verlag, 1991.

[17] R. Z. Khas'minskii. A limit theorem for the solutions of differential equations with random right-hand sides. *Theory Probab. Appl.*, 11:390–406, 1966.

[18] H. Kunita. *Stochastic Flows and Stochastic Differential Equations.* Cambridge University Press, 1990.

[19] T. G. Kurtz. A limit theorem for perturbed operator semigroups with applications to random evolutions. *J. Functional Anal.*, 12:55–67, 1973.

[20] T. G. Kurtz. Semigroups of conditioned shifts and approximation of Markov processes. *Ann. Probab.*, 3:618–642, 1975.

[21] T. G. Kurtz. Martingale problems for conditional distributions of Markov processes. *Elec. J. Probab.*, 3:1–29, 1998.

[22] T. G. Kurtz and D. L. Ocone. Unique characterization of conditional distributions in nonlinear filtering. *Ann. Probab.*, 16:80–107, 1988.

[23] T. G. Kurtz and R. H. Stockbridge. Existence of Markov controls and characterization of optimal Markov controls. *SIAM J. Control Optim.*, 36(2):609–653, 1998.

[24] H. J. Kushner. *Approximation and Weak Convergence Methods for Random Processes.* MIT Press, 1984.

[25] H. J. Kushner. *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems.* Birkhauser, 1990.

[26] H. J. Kushner. Nonlinear filtering for singularly perturbed systems. In *Stochastic Analysis (*E. Mayer-Wolf, E. Merzbach, and A. Shwartz, eds.*)*, pages 347–369. Academic Press, 1991.

[27] H. J. Kushner and H. Huang. Approximate and limit results for nonlinear filters with wide bandwidth observation noise. *Stochastics*, 16:65–96, 1986.

[28] V. M. Lucic and A. J. Heunis. On uniqueness of solutions for the stochastic differential equations of nonlinear filtering. *Ann. Appl. Probab. (to appear).*

[29] Yu. V. Prohorov. Convergence of random processes and limit theorems in probability theory. *Theory Probab. Appl.*, 1:157–214, 1956.

[30] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion.* Springer-Verlag, 1994.

[31] L. C. G. Rogers and D. Williams. *Diffusions, Markov Processes, and Martingales Vol. 2: Itô Calculus.* J. Wiley & Sons, 1987.

[32] L. C. G. Rogers and D. Williams. *Diffusions, Markov Processes, and Martingales Vol.1: Foundations.* J. Wiley & Sons, 1994.

[33] B. L. Rozovskii. A simple proof of uniqueness for Kushner and Zakai equations. In *Stochastic Analysis (*E. Mayer-Wolf, E. Merzbach, and A. Shwartz, eds.*)*, pages 449–458. Academic Press, 1991.

[34] R. L. Stratonovich. *Topics in the Theory of Random Noise, Vol. 1.* Gordon and Breach, 1963.

[35] D. W. Stroock and S. R. S. Varadhan. *Multidimensional Diffusion Processes.* Springer-Verlag, 1979.

[36] J. Szpirglas. Sur l'équivalence d'équations différentielles stochastiques á valeurs mesures intervenant dans le filtrage markovien non linéaire. *Ann. Inst. Henri Poincaré*, XIV(1):33–59, 1978.

[37] T. Yamada and S. Watanabe. On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.*, 11:155–167, 1971.

[38] M. Yor. Sur les théories du filtrage et de la prédiction. In *Séminaire de Probabilités XI, Lecture Notes in Mathematics, Vol. 581*, pages 257–297. Springer-Verlag, 1977.

[39] M. Zakai. On the optimal filtering of diffusion processes. *Z. Warsch. Verw. Gebiete*, 11:230–243, 1969.